Internet Engineering Task Force (IETF)
Request for Comments: 6703
Category: Informational
ISSN: 2070-1721

A. Morton
G. Ramachandran
G. Maguluri
AT&T Labs

August 2012

Reporting IP Network Performance Metrics: Different Points of View

Abstract

Consumers of IP network performance metrics have many different uses in mind. This memo provides "long-term" reporting considerations (e.g., hours, days, weeks, or months, as opposed to 10 seconds), based on analysis of the points of view of two key audiences. It describes how these audience categories affect the selection of metric parameters and options when seeking information that serves their needs.

Status of This Memo

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

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc6703.

Table of Contents

1. Introduction

When designing measurements of IP networks and presenting a result, knowledge of the audience is a key consideration. To present a useful and relevant portrait of network conditions, one must answer the following question:

"How will the results be used?"

There are two main audience categories for the report of results:

1. Network Characterization - describes conditions in an IP network for quality assurance, troubleshooting, modeling, Service Level Agreements (SLAs), etc. This point of view (POV) looks inward toward the network where the report consumer intends their actions.

2. Application Performance Estimation - describes the network conditions in a way that facilitates determining effects on user applications, and ultimately the users themselves. This POV looks outward, toward the user(s), accepting the network as is. This report consumer intends to estimate a network-dependent aspect of performance or design some aspect of an application's accommodation of the network. (These are *not* application metrics; they are defined at the IP layer.)

This memo considers how these different POVs affect both the measurement design (parameters and options of the metrics) and statistics reported when serving the report consumer's needs.

The IP Performance Metrics (IPPM) Framework [RFC2330] and other RFCs describing IPPM provide a background for this memo.

2. Purpose and Scope

The purpose of this memo is to clearly delineate two POVs for using measurements and describe their effects on the test design, including the selection of metric parameters and reporting the results.

The scope of this memo primarily covers the test design and reporting of the loss and delay metrics [RFC2680] [RFC2679]. It will also discuss the delay variation [RFC3393] and reordering metrics [RFC4737] where applicable.

With capacity metrics growing in relevance to the industry, the memo also covers POV and reporting considerations for metrics resulting from the Bulk Transfer Capacity Framework [RFC3148] and Network Capacity Definitions [RFC5136]. These memos effectively describe two different categories of metrics:

- Restricted [RFC3148]: includes restrictions of congestion control and the notion of unique data bits delivered, and
- Raw [RFC5136]: uses a definition of raw capacity without the restrictions of data uniqueness or congestion awareness.

It might seem, at first glance, that each of these metrics has an obvious audience (raw = network characterization, restricted = application performance), but reality is more complex and consistent with the overall topic of capacity measurement and reporting. For example, TCP is usually used in restricted capacity measurement methods, while UDP appears in raw capacity measurement. The raw and restricted capacity metrics will be treated in separate sections, although they share one common reporting issue: representing variability in capacity metric results as part of a long-term report.

Sampling, or the design of the active packet stream that is the basis for the measurements, is also discussed.

3. Reporting Results

This section gives an overview of recommendations, followed by additional considerations for reporting results in the "long term", based on the discussion and conclusions of the major sections that follow.

3.1. Overview of Metric Statistics

This section gives an overview of reporting recommendations for all the metrics considered in this memo.

The minimal report on measurements must include both loss and delay metrics.

For packet loss, the loss ratio defined in [RFC2680] is a sufficient starting point -- especially the existing guidance for setting the loss threshold waiting time. In Section 4.1.1, we have calculated a waiting time -- 51 seconds -- that should be sufficient to differentiate between packets that are truly lost or have long finite delays under general measurement circumstances. Knowledge of

specific conditions can help to reduce this threshold, and a waiting time of approximately 50 seconds is considered to be manageable in practice.

We note that a loss ratio calculated according to [Y.1540] would exclude errored packets from the numerator. In practice, the difference between these two loss metrics is small, if any, depending on whether the last link prior to the Destination contributes errored packets.

For packet delay, we recommend providing both the mean delay and the median delay with lost packets designated as undefined (as permitted by [RFC2679]). Both statistics are based on a conditional distribution, and the condition is packet arrival prior to a waiting time dT, where dT has been set to take maximum packet lifetimes into account, as discussed above for loss. Using a long dT helps to ensure that delay distributions are not truncated.

For Packet Delay Variation (PDV), the minimum delay of the conditional distribution should be used as the reference delay for computing PDV according to [Y.1540] or [RFC5481] and [RFC3393]. A useful value to report is a "pseudo" range of delay variation based on calculating the difference between a high percentile of delay and the minimum delay. For example, the 99.9th percentile minus the minimum will give a value that can be compared with objectives in [Y.1541].

For both raw capacity and restricted capacity, reporting the variability in a useful way is identified as the main challenge. The min, max, and range statistics are suggested along with a ratio of max to min and moving averages. In the end, a simple plot of the singleton results over time may succeed where summary metrics fail or may serve to confirm that the summaries are valid.

3.2. Long-Term Reporting Considerations

[IPPM-RPT] describes methods to conduct measurements and report the results on a near-immediate time scale (10 seconds, which we consider to be "short-term").

Measurement intervals and reporting intervals need not be the same length. Sometimes, the user is only concerned with the performance levels achieved over a relatively long interval of time (e.g., days, weeks, or months, as opposed to 10 seconds). However, there can be risks involved with running a measurement continuously over a long period without recording intermediate results:

- Temporary power failure may cause loss of all results to date.
- Measurement system timing synchronization signals may experience a temporary outage, causing subsets of measurements to be in error or invalid.
- Maintenance on the measurement system or on its connectivity to the network under test may be necessary.

For these and other reasons, such as

- the constraint to collect measurements on intervals similar to user session length,
- the dual use of measurements in monitoring activities where results are needed on a period of a few minutes, or
- the ability to inspect results of a single measurement interval for deeper analysis,

there is value in conducting measurements on intervals that are much shorter than the reporting interval.

There are several approaches for aggregating a series of measurement results over time in order to make a statement about the longer reporting interval. One approach requires the storage of all metric singletons collected throughout the reporting interval, even though the measurement interval stops and starts many times.

Another approach is described in [RFC5835] as "temporal aggregation". This approach would estimate the results for the reporting interval based on combining many individual short-term measurement interval statistics to yield a long-term result. The result would ideally appear in the same form as though a continuous measurement had been conducted. A memo addressing the details of temporal aggregation is yet to be prepared.

Yet another approach requires a numerical objective for the metric, and the results of each measurement interval are compared with the objective. Every measurement interval where the results meet the objective contribute to the fraction of time with performance as specified. When the reporting interval contains many measurement intervals, it is possible to present the results as "metric A was less than or equal to objective X during Y% of time".

> NOTE that numerical thresholds of acceptability are not set in IETF performance work and are therefore excluded from the scope of this memo.

In all measurements, it is important to avoid unintended synchronization with network events. This topic is treated in [RFC2330] for Poisson-distributed inter-packet time streams and in [RFC3432] for Periodic streams. Both avoid synchronization by using random start times.

There are network conditions where it is simply more useful to report the connectivity status of the Source-Destination path, and to distinguish time intervals where connectivity can be demonstrated from other time intervals (where connectivity does not appear to exist). [RFC2678] specifies a number of one-way and two-way connectivity metrics of increasing complexity. In this memo, we recommend that long-term reporting of loss, delay, and other metrics be limited to time intervals where connectivity can be demonstrated, and that other intervals be summarized as the percent of time where connectivity does not appear to exist. We note that this same approach has been adopted in ITU-T Recommendation [Y.1540] where performance parameters are only valid during periods of service "availability" (evaluated according to a function based on packet loss, and sustained periods of loss ratio greater than a threshold are declared "unavailable").

4. Effect of POV on the Loss Metric

This section describes the ways in which the loss metric can be tuned to reflect the preferences of the two audience categories, or different POVs. The waiting time before declaring that a packet is lost -- the loss threshold -- is one area where there would appear to be a difference, but the ability to post-process the results may resolve it.

4.1. Loss Threshold

RFC 2680 [RFC2680] defines the concept of a waiting time for packets to arrive, beyond which they are declared lost. The text of the RFC declines to recommend a value, instead saying that "good engineering, including an understanding of packet lifetimes, will be needed in practice". Later, in the methodology, they give reasons for waiting "a reasonable period of time" and leave the definition of "reasonable" intentionally vague. Below, we estimate a practical bound on waiting time.

4.1.1. Network Characterization

Practical measurement experience has shown that unusual network circumstances can cause long delays. One such circumstance is when routing loops form during IGP re-convergence following a failure or drastic link cost change. Packets will loop between two routers

until new routes are installed or until the IPv4 Time-to-Live (TTL) field (or the IPv6 Hop Limit) decrements to zero. Very long delays on the order of several seconds have been measured [Casner] [Cia03].

Therefore, network characterization activities prefer a long waiting time in order to distinguish these events from other causes of loss (such as packet discard at a full queue, or tail drop). This way, the metric design helps to distinguish more reliably between packets that might yet arrive and those that are no longer traversing the network.

It is possible to calculate a worst-case waiting time, assuming that a routing loop is the cause. We model the path between Source and Destination as a series of delays in links (t) and queues (q), as these are the dominant contributors to delay (in active measurement, the Source and Destination hosts contribute minimal delay). The normal path delay, D, across n queues (where TTL is decremented at a node with a queue) and n+1 links without encountering a loop, is

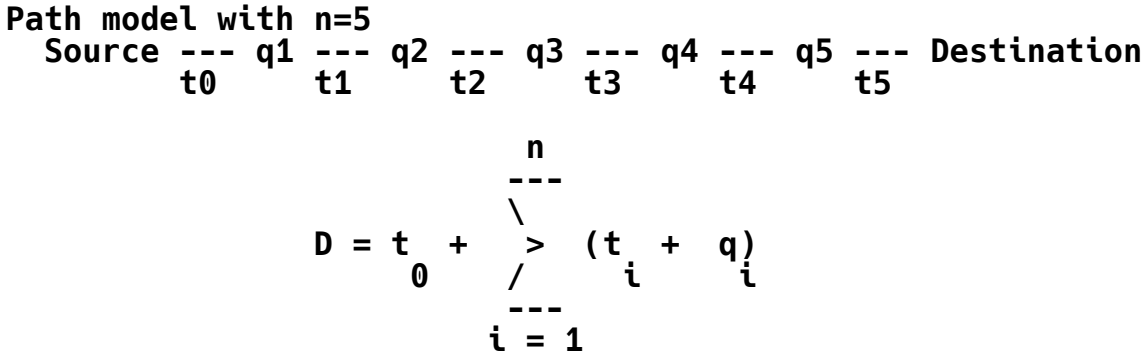

Figure 1: Normal Path Delay

and the time spent in the loop with L queues is

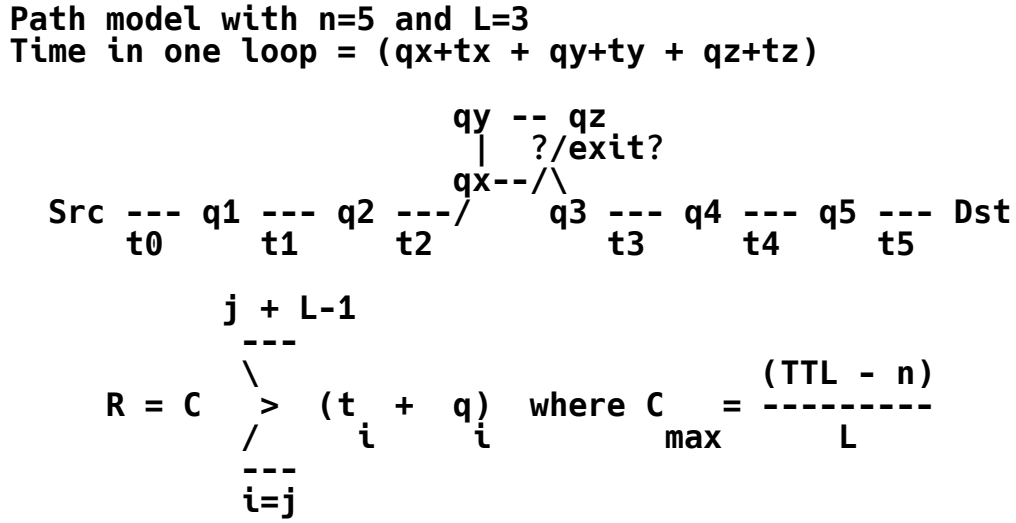

Figure 2: Delay Due to Rotations in a Loop

where n is the total number of queues in the non-loop path (with n+1 links), j is the queue number where the loop begins, C is the number of times a packet circles the loop, and TTL is the packet's initial Time-to-Live value at the Source (or Hop Count in IPv6).

If we take the delays of all links and queues as 100 ms each, the TTL=255, the number of queues n=5, and the queues in the loop L=4, then using C_max:

D = 1.1 seconds and R ~= 50 seconds, and D + R ~= 51.1 seconds

We note that the link delays of 100 ms would span most continents, and a constant queue length of 100 ms is also very generous. When a loop occurs, it is almost certain to be resolved in 10 seconds or less. The value calculated above is an upper limit for almost any real-world circumstance.

A waiting time threshold parameter, dT, set consistent with this calculation, would not truncate the delay distribution (possibly causing a change in its mathematical properties), because the packets that might arrive have been given sufficient time to traverse the network.

It is worth noting that packets that are stored and deliberately forwarded at a much later time constitute a replay attack on the measurement system and are beyond the scope of normal performance reporting.

4.1.2. Application Performance

Fortunately, application performance estimation activities are not adversely affected by the long estimated limit on waiting time, because most applications will use shorter time thresholds. Although the designer's tendency might be to set the loss threshold at a value equivalent to a particular application's threshold, this specific threshold can be applied when post-processing the measurements. A shorter waiting time can be enforced by locating packets with delays longer than the application's threshold and re-designating such packets as lost. Thus, the measurement system can use a single loss waiting time and support both application and network performance POVs simultaneously.

4.2. Errored Packet Designation

RFC 2680 designates packets that arrive containing errors as lost packets. Many packets that are corrupted by bit errors are discarded within the network and do not reach their intended destination.

This is consistent with applications that would check the payload integrity at higher layers and discard the packet. However, some applications prefer to deal with errored payloads on their own, and even a corrupted payload is better than no packet at all.

To address this possibility, and to make network characterization more complete, distinguishing between packets that do not arrive (lost) and errored packets that arrive (conditionally lost) is recommended.

4.3. Causes of Lost Packets

Although many measurement systems use a waiting time to determine whether or not a packet is lost, most of the waiting is in vain. The packets are no longer traversing the network and have not reached their destination.

There are many causes of packet loss, including the following:

1. Queue drop, or discard

2. Corruption of the IP header, or other essential header information

3. TTL expiration (or use of a TTL value that is too small)

4. Link or router failure

5. Layers below the Source-to-Destination IP layer can discard packets that fail error checking, and link-layer checksums often cover the entire packet

It is reasonable to consider a packet that has not arrived after a large amount of time to be lost (due to one of the causes above) because packets do not "live forever" in the network or have infinite delay.

4.4. Summary for Loss

Given that measurement post-processing is possible (even encouraged in the definitions of IPPM), measurements of loss can easily serve both POVs:

- Use a long waiting time to serve network characterization and revise results for specific application delay thresholds as needed.

- Distinguish between errored packets and lost packets when possible to aid network characterization, and combine the results for application performance if appropriate.

5. Effect of POV on the Delay Metric

This section describes the ways in which the delay metric can be tuned to reflect the preferences of the two consumer categories, or different POVs.

5.1. Treatment of Lost Packets

The delay metric [RFC2679] specifies the treatment of packets that do not successfully traverse the network: their delay is undefined.

> >>The *Type-P-One-way-Delay* from Src to Dst at T is undefined (informally, infinite)<< means that Src sent the first bit of a Type-P packet to Dst at wire-time T and that Dst did not receive that packet.

It is an accepted but informal practice to assign infinite delay to lost packets. We next look at how these two different treatments align with the needs of measurement consumers who wish to characterize networks or estimate application performance. Also, we look at the way that lost packets have been treated in other metrics: delay variation and reordering.

5.1.1. Application Performance

Applications need to perform different functions, dependent on whether or not each packet arrives within some finite tolerance. In other words, a receiver's packet processing takes only one of two alternative directions (a "fork" in the road):

- Packets that arrive within expected tolerance are handled by removing headers, restoring smooth delivery timing (as in a de-jitter buffer), restoring sending order, checking for errors in payloads, and many other operations.

- Packets that do not arrive when expected lead to attempted recovery from the apparent loss, such as retransmission requests, loss concealment, or forward error correction to replace the missing packet.

So, it is important to maintain a distinction between packets that actually arrive and those that do not. Therefore, it is preferable to leave the delay of lost packets undefined and to characterize the delay distribution as a conditional distribution (conditioned on arrival).

5.1.2. Network Characterization

In this discussion, we assume that both loss and delay metrics will be reported for network characterization (at least).

Assume that packets that do not arrive are reported as lost, usually as a fraction of all sent packets. If these lost packets are assigned an undefined delay, then the network's inability to deliver them (in a timely way) is relegated only in the loss metric when we report statistics on the delay distribution conditioned on the event of packet arrival (within the loss waiting time threshold). We can say that the delay and loss metrics are orthogonal in that they convey non-overlapping information about the network under test. This is a valuable property whose absence is discussed below.

However, if we assign infinite delay to all lost packets, then

- The delay metric results are influenced both by packets that arrive and those that do not.

- The delay singleton and the loss singleton do not appear to be orthogonal (delay is finite when loss=0; delay is infinite when loss=1).

o The network is penalized in both the loss and delay metrics, effectively double-counting the lost packets.

As further evidence of overlap, consider the Cumulative Distribution Function (CDF) of delay when the value "positive infinity" is assigned to all lost packets. Figure 3 shows a CDF where a small fraction of packets are lost.

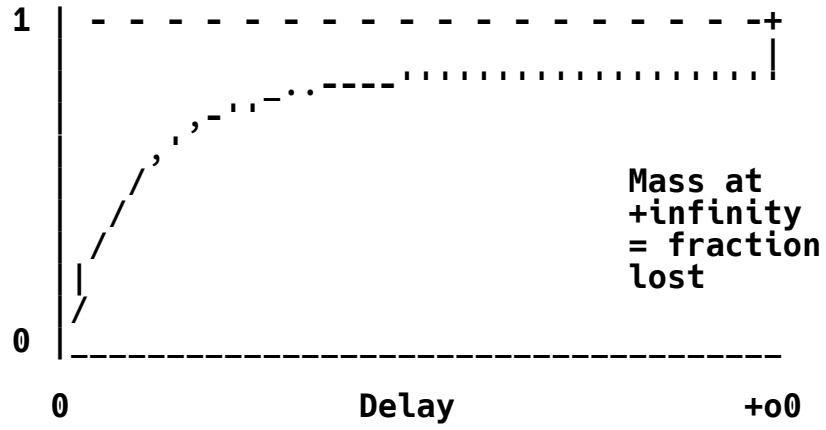

Figure 3: Cumulative Distribution Function for Delay When Loss = +Infinity

We note that a delay CDF that is conditioned on packet arrival would not exhibit this apparent overlap with loss.

Although infinity is a familiar mathematical concept, it is somewhat disconcerting to see any time-related metric reported as infinity. Questions are bound to arise and tend to detract from the goal of informing the consumer with a performance report.

5.1.3. Delay Variation

[RFC3393] excludes lost packets from samples, effectively assigning an undefined delay to packets that do not arrive in a reasonable time. Section 4.1 of [RFC3393] describes this specification and its rationale (ipdv = inter-packet delay variation in the quote below).

> The treatment of lost packets as having "infinite" or "undefined" delay complicates the derivation of statistics for ipdv. Specifically, when packets in the measurement sequence are lost, simple statistics such as sample mean cannot be computed. One possible approach to handling this problem is to reduce the event space by conditioning. That is, we consider conditional statistics; namely we estimate the mean ipdv (or other derivative statistic) conditioned on the event that selected packet pairs

arrive at the Destination (within the given timeout). While this itself is not without problems (what happens, for example, when every other packet is lost), it offers a way to make some (valid) statements about ipdv, at the same time avoiding events with undefined outcomes.

We note that the argument above applies to all forms of packet delay variation that can be constructed using the "selection function" concept of [RFC3393]. In recent work, the two main forms of delay variation metrics have been compared, and the results are summarized in [RFC5481].

5.1.4. Reordering

[RFC4737] defines metrics that are based on evaluation of packet arrival order and that include a waiting time before declaring that a packet is lost (to exclude the packet from further processing).

If packets are assigned a delay value, then the reordering metric would declare any packets with infinite delay to be reordered, because their sequence numbers will surely be less than the "Next Expected" threshold when (or if) they arrive. But this practice would fail to maintain orthogonality between the reordering metric and the loss metric. Confusion can be avoided by designating the delay of non-arriving packets as undefined and reserving delay values only for packets that arrive within a sufficiently long waiting time.

5.2. Preferred Statistics

Today in network characterization, the sample mean is one statistic that is almost ubiquitously reported. It is easily computed and understood by virtually everyone in this audience category. Also, the sample is usually filtered on packet arrival, so that the mean is based on a conditional distribution.

The median is another statistic that summarizes a distribution, having somewhat different properties from the sample mean. The median is stable in distributions with a few outliers or without them. However, the median's stability prevents it from indicating when a large fraction of the distribution changes value. 50% or more values would need to change for the median to capture the change.

Both the median and sample mean have difficulty with bimodal distributions. The median will reside in only one of the modes, and the mean may not lie in either mode range. For this and other reasons, additional statistics such as the minimum, maximum, and 95th percentile have value when summarizing a distribution.

When both the sample mean and median are available, a comparison will sometimes be informative, because these two statistics are equal only under unusual circumstances, such as when the delay distribution is perfectly symmetrical.

Also, these statistics are generally useful from the application performance POV, so there is a common set that should satisfy audiences.

Plots of the delay distribution may also be useful when single-value statistics indicate that new conditions are present. An empirically derived probability distribution function will usually describe multiple modes more efficiently than any other form of result.

5.3. Summary for Delay

From the perspectives of

1. application/receiver analysis, where subsequent processing depends on whether the packet arrives or times out,

2. straightforward network characterization without double-counting defects, and

3. consistency with delay variation and reordering metric definitions,

the most efficient practice is to distinguish between packets that are truly lost and those that are delayed packets with a sufficiently long waiting time, and to designate the delay of non-arriving packets as undefined.

6. Reporting Raw Capacity Metrics

Raw capacity refers to the metrics defined in [RFC5136], which do not include restrictions such as data uniqueness or flow-control response to congestion.

The metrics considered are IP-layer capacity, utilization (or used capacity), and available capacity, for individual links and complete paths. These three metrics form a triad: knowing one metric constrains the other two (within their allowed range), and knowing two determines the third. The link metrics have another key aspect in common: they are single-measurement-point metrics at the egress of a link. The path capacity and available capacity are derived by examining the set of single-point link measurements and taking the minimum value.

6.1. Type-P Parameter

The concept of "packets of Type-P" is defined in [RFC2330]. The Type-P categorization has critical relevance in all forms of capacity measurement and reporting. The ability to categorize packets based on header fields for assignment to different queues and scheduling mechanisms is now commonplace. When unused resources are shared across queues, the conditions in all packet categories will affect capacity and related measurements. This is one source of variability in the results that all audiences would prefer to see reported in a useful and easily understood way.

Communication of Type-P within the One-Way Active Measurement Protocol (OWAMP) and the Two-Way Active Measurement Protocol (TWAMP) is essentially confined to the Diffserv Code Point (DSCP) [RFC4656]. DSCP is the most common qualifier for Type-P.

Each audience will have a set of Type-P qualifications and value combinations that are of interest. Measurements and reports should have the flexibility to report per-type and aggregate performance.

6.2. A priori Factors

The audience for network characterization may have detailed information about each link that comprises a complete path (due to ownership, for example), or some of the links in the path but not others, or none of the links.

There are cases where the measurement audience only has information on one of the links (the local access link) and wishes to measure one or more of the raw capacity metrics. This scenario is quite common and has spawned a substantial number of experimental measurement methods (e.g., http://www.caida.org/tools/taxonomy/). Many of these methods respect that their users want a result fairly quickly and in one trial. Thus, the measurement interval is kept short (a few seconds to a minute). For long-term reporting, a sample of short-term results needs to be summarized.

6.3. IP-Layer Capacity

For links, this metric's theoretical maximum value can be determined from the physical-layer bit rate and the bit rate reduction due to the layers between the physical layer and IP. When measured, this metric takes additional factors into account, such as the ability of the sending device to process and forward traffic under various conditions. For example, the arrival of routing updates may spawn high-priority processes that reduce the sending rate temporarily.

Thus, the measured capacity of a link will be variable, and the maximum capacity observed applies to a specific time, time interval, and other relevant circumstances.

For paths composed of a series of links, it is easy to see how the sources of variability for the results grow with each link in the path. Variability of results will be discussed in more detail below.

6.4. IP-Layer Utilization

The ideal metric definition of link utilization [RFC5136] is based on the actual usage (bits successfully received during a time interval) and the maximum capacity for the same interval.

In practice, link utilization can be calculated by counting the IP-layer (or other layer) octets received over a time interval and dividing by the theoretical maximum number of octets that could have been delivered in the same interval. A commonly used time interval is 5 minutes, and this interval has been sufficient to support network operations and design for some time. 5 minutes is somewhat long compared with the expected download time for web pages but short with respect to large file transfers and TV program viewing. It is fair to say that considerable variability is concealed by reporting a single (average) utilization value for each 5-minute interval. Some performance management systems have begun to make 1-minute averages available.

There is also a limit on the smallest useful measurement interval. Intervals on the order of the serialization time for a single Maximum Transmission Unit (MTU) packet will observe on/off behavior and report 100% or 0%. The smallest interval needs to be some multiple of MTU serialization time for averaging to be effective.

6.5. IP-Layer Available Capacity

The available capacity of a link can be calculated using the capacity and utilization metrics.

When available capacity of a link or path is estimated through some measurement technique, the following parameters should be reported:

- Name and reference to the exact method of measurement
- IP packet length, octets (including IP header)
- Maximum capacity that can be assessed in the measurement configuration

- Time duration of the measurement
- All other parameters specific to the measurement method

Many methods of available capacity measurement have a maximum capacity that they can measure, and this maximum may be less than the actual available capacity of the link or path. Therefore, it is important to know the capacity value beyond which there will be no measured improvement.

The application performance estimation audience may have a desired target capacity value and simply wish to assess whether there is sufficient available capacity. This case simplifies the measurement of link and path capacity to some degree, as long as the measurable maximum exceeds the target capacity.

6.6. Variability in Utilization and Available Capacity

As with most metrics and measurements, assessing the consistency or variability in the results gives the user an intuitive feel for the degree (or confidence) that any one value is representative of other results, or the spread of the underlying distribution of the singleton measurements.

How can utilization be measured and summarized to describe the potential variability in a useful way?

How can the variability in available capacity estimates be reported, so that the confidence in the results is also conveyed?

We suggest some methods below.

6.6.1. General Summary of Variability

With a set of singleton utilization or available capacity estimates, each representing a time interval needed to ascertain the estimate, we seek to describe the variation over the set of singletons as though reporting summary statistics of a distribution. Three useful summary statistics are

- Minimum,
- Maximum, and
- Range

An alternate way to represent the range is as a ratio of maximum to minimum value. This enables an easily understandable statistic to describe the range observed. For example, when maximum = 3*minimum, then the max/min ratio is 3, and users may see variability of this order. On the other hand, capacity estimates with a max/min ratio near 1 are quite consistent and near the central measure or statistic reported.

For an ongoing series of singleton estimates, a moving average of n estimates may provide a single value estimate to more easily distinguish substantial changes in performance over time. For example, in a window of n singletons observed in time interval t, a percentage change of x% is declared to be a substantial change and reported as an exception.

Often, the most informative summary of the results is a two-axis plot rather than a table of statistics, where time is plotted on the x-axis and the singleton value on the y-axis. The time-series plot can illustrate sudden changes in an otherwise stable range, identify bi-modality easily, and help quickly assess correlation with other time-series. Plots of frequency of the singleton values are likewise useful tools to visualize the variation.

7. Reporting Restricted Capacity Metrics

Restricted capacity refers to the metrics defined in [RFC3148], which include criteria of data uniqueness or flow-control response to congestion.

One primary metric considered is Bulk Transfer Capacity (BTC) for complete paths. [RFC3148] defines BTC as

```
BTC = data_sent / elapsed_time
```

for a connection with congestion-aware flow control, where data_sent is the total number of unique payload bits (no headers).

We note that this definition *differs* from the raw capacity definition in Section 2.3.1 of [RFC5136], where IP-layer capacity *includes* all bits in the IP header and payload. This means that restricted capacity BTC is already operating at a disadvantage when compared to the raw capacity at layers below TCP. Further, there are cases where one IP layer is encapsulated in another IP layer or other form of tunneling protocol, designating more and more of the fundamental transport capacity as header bits that are pure overhead to the BTC measurement.

We also note that raw and restricted capacity metrics are not orthogonal in the sense defined in Section 5.1.2 above. The information they convey about the network under test is certainly overlapping, but they reveal two different and important aspects of performance.

When thinking about the triad of raw capacity metrics, BTC is most akin to the "IP-Type-P Available Path Capacity", at least in the eyes of a network user who seeks to know what transmission performance a path might support.

7.1. Type-P Parameter and Type-C Parameter

The concept of "packets of Type-P" is defined in [RFC2330]. The considerations for restricted capacity are identical to the raw capacity section on this topic, with the addition that the various fields and options in the TCP header must be included in the description.

The vast array of TCP flow-control options are not well captured by Type-P, because they do not exist in the TCP header bits. Therefore, we introduce a new notion here: TCP Configuration of "Type-C". The elements of Type-C describe all of the settings for TCP options and congestion control algorithm variables, including the main form of congestion control in use. Readers should consider the parameters and variables of [RFC3148] and [RFC6349] when constructing Type-C.

7.2. A Priori Factors

The audience for network characterization may have detailed information about each link that comprises a complete path (due to ownership, for example), or some of the links in the path but not others, or none of the links.

There are cases where the measurement audience only has information on one of the links (the local access link) and wishes to measure one or more BTC metrics. The discussion in Section 6.2 applies here as well.

7.3. Measurement Interval

There are limits on a useful measurement interval for BTC. Three factors that influence the interval duration are listed below:

1. Measurements may choose to include or exclude the 3-way handshake of TCP connection establishment, which requires at least 1.5 * RTT (round-trip time) and contains both the delay of the path and the host processing time for responses. However, user experience includes the 3-way handshake for all new TCP connections.

2. Measurements may choose to include or exclude Slow-Start, preferring instead to focus on a portion of the transfer that represents "equilibrium" (which needs to be defined for particular circumstances if used). However, user experience includes the Slow-Start for all new TCP connections.

3. Measurements may choose to use a fixed block of data to transfer, where the size of the block has a relationship to the file size of the application of interest. This approach yields variable size measurement intervals, where a path with faster BTC is measured for less time than a path with slower BTC, and this has implications when path impairments are time-varying, or transient. Users are likely to turn their immediate attention elsewhere when a very large file must be transferred; thus, they do not directly experience such a long transfer -- they see the result (success or failure) and possibly an objective measurement of the transfer time (which will likely include the 3-way handshake, Slow-Start, and application file management processing time as well as the BTC).

Individual measurement intervals may be short or long, but there is a need to report the results on a long-term basis that captures the BTC variability experienced between each interval. Consistent BTC is a valuable commodity along with the value attained.

7.4. Bulk Transfer Capacity Reporting

When BTC of a link or path is estimated through some measurement technique, the following parameters should be reported:

- Name and reference to the exact method of measurement
- Maximum Transmission Unit (MTU)
- Maximum BTC that can be assessed in the measurement configuration
- Time and duration of the measurement

- Number of BTC connections used simultaneously
- *All* other parameters specific to the measurement method, especially the congestion control algorithm in use

See also [RFC6349].

Many methods of BTC measurement have a maximum capacity that they can measure, and this maximum may be less than the available capacity of the link or path. Therefore, it is important to specify the measured BTC value beyond which there will be no measured improvement.

The application performance estimation audience may have a desired target capacity value and simply wish to assess whether there is sufficient BTC. This case simplifies the measurement of link and path capacity to some degree, as long as the measurable maximum exceeds the target capacity.

7.5. Variability in Bulk Transfer Capacity

As with most metrics and measurements, assessing the consistency or variability in the results gives the user an intuitive feel for the degree (or confidence) that any one value is representative of other results, or the underlying distribution from which these singleton measurements have come.

With two questions looming --

1. What ways can BTC be measured and summarized to describe the potential variability in a useful way?
2. How can the variability in BTC estimates be reported, so that the confidence in the results is also conveyed?

-- we suggest the methods listed in Section 6.6.1 above, and the additional results presentations given in [RFC6349].

8. Reporting on Test Streams and Sample Size

This section discusses two key aspects of measurement that are sometimes omitted from the report: the description of the test stream on which the measurements are based, and the sample size.

8.1. Test Stream Characteristics

Network characterization has traditionally used Poisson-distributed inter-packet spacing, as this provides an unbiased sample. The average inter-packet spacing may be selected to allow observation of

specific network phenomena. Other test streams are designed to sample some property of the network, such as the presence of congestion, link bandwidth, or packet reordering.

If measuring a network in order to make inferences about applications or receiver performance, then there are usually efficiencies derived from a test stream that has similar characteristics to the sender. In some cases, it is essential to synthesize the sender stream, as with BTC estimates. In other cases, it may be sufficient to sample with a "known bias", e.g., a Periodic stream to estimate real-time application performance.

8.2. Sample Size

Sample size is directly related to the accuracy of the results and plays a critical role in the report. Even if only the sample size (in terms of number of packets) is given for each value or summary statistic, it imparts a notion of the confidence in the result.

In practice, the sample size will be selected taking both statistical and practical factors into account. Among these factors are the following:

1. The estimated variability of the quantity being measured.

2. The desired confidence in the result (although this may be dependent on assumption of the underlying distribution of the measured quantity).

3. The effects of active measurement traffic on user traffic.

A sample size may sometimes be referred to as "large". This is a relative and qualitative term. It is preferable to describe what one is attempting to achieve with his sample. For example, stating an implication may be helpful: this sample is large enough that a single outlying value at ten times the "typical" sample mean (the mean without the outlying value) would influence the mean by no more than X.

The Appendix of [RFC2330] indicates that a sample size of 128 singletons worked well for goodness-of-fit testing, while a much larger size (8192 singletons) almost always failed.

9. Security Considerations

The security considerations that apply to any active measurement of live networks are relevant here as well. See the Security Considerations section of [RFC4656] for mandatory-to-implement security features that intend to mitigate attacks.

Measurement systems conducting long-term measurements are more exposed to threats as a by-product of ports open longer to perform their task, and more easily detected measurement activity on those ports. Further, use of long packet waiting times affords an attacker a better opportunity to prepare and launch a replay attack.

10. Acknowledgements

The authors thank Phil Chimento for his suggestion to employ conditional distributions for delay, Steve Konish Jr. for his careful review and suggestions, Dave McDysan and Don McLachlan for useful comments based on their long experience with measurement and reporting, Daniel Genin for his observation of non-orthogonality between raw and restricted capacity metrics (and for noticing our previous omission of this fact), and Matt Zekauskas for suggestions on organizing the memo for easier consumption.

11. References

11.1. Normative References

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics", RFC 2330, May 1998.

[RFC2678] Mahdavi, J. and V. Paxson, "IPPM Metrics for Measuring Connectivity", RFC 2678, September 1999.

[RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Delay Metric for IPPM", RFC 2679, September 1999.

[RFC2680] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way Packet Loss Metric for IPPM", RFC 2680, September 1999.

[RFC3148] Mathis, M. and M. Allman, "A Framework for Defining Empirical Bulk Transfer Capacity Metrics", RFC 3148, July 2001.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, November 2002.

[RFC3432] Raisanen, V., Grotefeld, G., and A. Morton, "Network performance measurement with periodic streams", RFC 3432, November 2002.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", RFC 4656, September 2006.

[RFC4737] Morton, A., Ciavattone, L., Ramachandran, G., Shalunov, S., and J. Perser, "Packet Reordering Metrics", RFC 4737, November 2006.

[RFC5136] Chimento, P. and J. Ishac, "Defining Network Capacity", RFC 5136, February 2008.

11.2. Informative References

[Casner] Casner, S., Alaettinoglu, C., and C. Kuan, "A Fine-Grained View of High-Performance Networking", NANOG 22 Conf., May 20-22 2001, <http://www.nanog.org/presentations/archive/index.php>.

[Cia03] Ciavattone, L., Morton, A., and G. Ramachandran, "Standardized Active Measurements on a Tier 1 IP Backbone", IEEE Communications Magazine, Vol. 41 No. 6, pp. 90-97, June 2003.

[IPPM-RPT] Shalunov, S. and M. Swany, "Reporting IP Performance Metrics to Users", Work in Progress, March 2011.

[RFC5481] Morton, A. and B. Claise, "Packet Delay Variation Applicability Statement", RFC 5481, March 2009.

[RFC5835] Morton, A., Ed., and S. Van den Berghe, Ed., "Framework for Metric Composition", RFC 5835, April 2010.

[RFC6349] Constantine, B., Forget, G., Geib, R., and R. Schrage, "Framework for TCP Throughput Testing", RFC 6349, August 2011.

[Y.1540] International Telecommunication Union, "Internet protocol data communication service - IP packet transfer and availability performance parameters", ITU-T Recommendation Y.1540, March 2011.

[Y.1541] International Telecommunication Union, "Network performance objectives for IP-based services", ITU-T Recommendation Y.1541, December 2011.

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: acmorton@att.com
URI: http://home.comcast.net/~acmacm/

Gomathi Ramachandran
AT&T Labs
200 Laurel Avenue South
Middletown, New Jersey 07748
USA

Phone: +1 732 420 2353
EMail: gomathi@att.com

Ganga Maguluri
AT&T Labs
200 Laurel Avenue South
Middletown, New Jersey 07748
USA

Phone: +1 732 420 2486
EMail: gmaguluri@att.com