ALTO Performance Metrics

draft-ietf-alto-performance-metrics-12

Luis Miguel Contreras Murillo Sabine Randriamasy Y. Richard Yang

> D. Dhody Young Lee Qin Wu

July 27, 2020

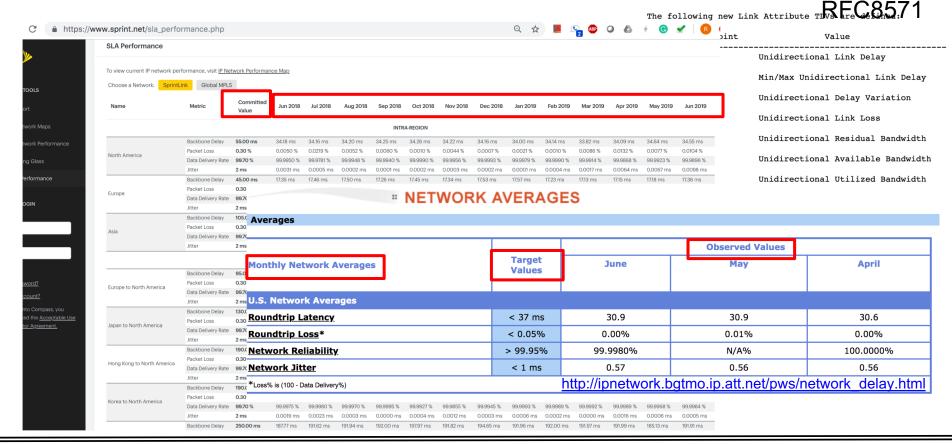
IETF 108, Virtual Meeting

Outline

- Updates from v10-v12: address discussions and reviews at interim meeting in April 2020
 - How to choose types of metrics
 - Conforming to RFC6390 (how much details to specify)
 - How to handle different statistics of the same metric
 - How to convey freshness of metric values

Recall of Key WG Decision

- ALTO provides guidance, not measurement framework
- There can be multiple types of guidance; we chose 4 types, but how to specify the type



Import vs estimation

Section 2.1:

- The "estimation" category indicates that the value of the metric is computed through an estimation process. An ALTO server may compute "estimation" values by retrieving and/or aggregating information from routing protocols (e.g., [RFC8571]) and traffic measurement management tools (e.g., TWAMP [RFC5357]), with corresponding operational issues.
- A particular type of "estimation is direct "import", which indicates that the value of the metric is imported directly from a specific existing protocol or system. Specifying "import" as source instead of the more generic "estimation" may allow better tracing of information flow. For an "import" metric, it is RECOMMENDED that the "parameters" field provides details to the system from which raw data is imported. In particular, one may notice that the set of end-to-end metrics defined in Table 1 has large overlap with the set defined in [RFC8571], in the setting of IGP traffic engineering performance metrics for each link (i.e., unidirectional link delay, min/max unidirectional link delay, unidirectional delay variation, unidirectional link loss, unidirectional residual bandwidth, unidirectional available bandwidth, unidirectional utilized bandwidth). Hence, an ALTO server may use "import" to indicate that its end-to-end metrics are computed from link metrics imported from [RFC8571].
- There can be overlap in deciding the cost-source category. It is the operator of an ALTO server who chooses the category. If a metric does not include a "cost-source" value, the application MUST assume that the value of "cost-source" is the most generic "estimation".

Conforming to [RFC6390]

Section 2.

- "When defining the metrics in Table 1, this document considers the guidelines specified in [RFC6390], which requires fine-grained specification of (i) Metric Name, (ii) Metric Description, (iii) Method of Measurement or Calculation, (iv) Units of Measurement, (v) Measurement Points, and (vi) Measurement Timing. In particular, for each metric, this document defines (i) Metric Name, (ii) Metric Description, and (iv) Units of Measurement. The Measurement Points are always specified by the specific ALTO services; for example, endpoint cost service is between the two end points.

On the other hand, to be able to use coarse-grained information such as routing system information (e.g., [RFC8571]), which may not provide fine-grained information such as (iii) Method of Measurement or Calculation and (vi) Measurement Timing, this document provides context information to indicate the source of information and hence available metric details."

How to Handle Statistics of Same Metric

- "Q2: How to handle statistics of the same metric
 - statistics: min, max, x-percentile, avg, ...
 - related to Martin Duke comment"
- Updated Sec. 2.2:
 - "The measurement of a performance metric often yields a set of samples from an observation distribution ([Prometheus]), instead of a single value. This document considers that the samples are aggregated as a statistic when reported. Hence, each performance metric's identifier should indicate the statistic (i.e., an aggregation operation), to become <metric-base-identifier>-<stat>"
 - Compared with other common statistics such as those in Prometheus [1]
 - They use quantile but we decide to keep percentile
 - Make complete statistics: add stddev, stdvar
 percentile, min, max, median, mean, stddev, stdvar
 - [1] https://prometheus.io/docs/concepts/data_model/

Freshness of Metrics

```
5. Operational Considerations
                                              4.3.3. Intended Semantics and Use . . . . .
    4.3.4. Cost-Context Specification Considerations
    5. Operational Considerations . . . . . . . . . .
 Metric Timestamp Consideration . . . . . . . .
  5.3.1. Configuration Parameters Considerations . . . . . .
  5.3.2. Availability Considerations . . . . . . . . . . . . . . . . .
                                                Backward Compatibility Considerations . . . . .
5.4.1. Configuration Parameters Considerations . .
                                                  Availability Considerations . . .
```

"Despite the introduction of the additional cost-context information, there is not a built-in field to indicate the timestamps of the data used to compute a metric. To indicate this attribute, the ALTO server SHOULD return HTTP "Last-Modified", to indicate the freshness of the data used to compute the performance metrics. If the ALTO client obtains updates through an incremental update mechanism (e.g., RFC editor: Fix the RFC number when available. [ALTO SSE]), the client SHOULD assume that the metric is computed using a snapshot at the time that is approximated by the receiving time."

Next Step

- The authors are quite happy about the documents, with some very minor edits needed, e.g.,
 - Adding Content-Length in the examples after no more changes