

Precision Availability Metrics for SLO-Governed End-to-end Services

draft-mhmcsfh-ippm-pam

Greg Mirsky
Joel Halpern
Xiao Min
Liuyan Han
Adrian Farrel

Alexander Clemm
John Strassner
Jérôme François
Mohamed Boucadair

IETF-115, November 2022

- What is the status of this work?

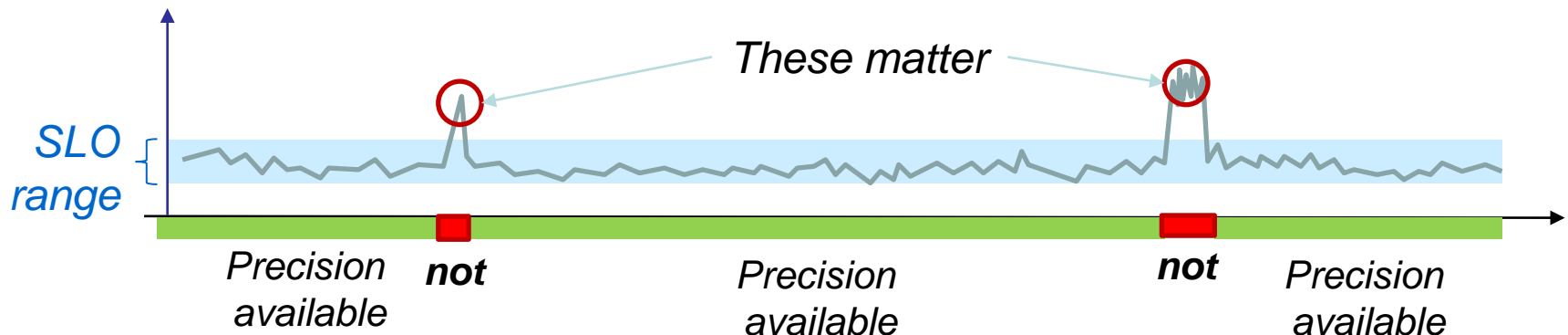
The draft is the result of merging work of two teams. The draft is in the WG AP at the IPPM WG.

- Why are we presenting this to TEAS?

Because it is important for traffic engineering and network slicing operations, and TEAS has the expertise on those topics.

What Is the Problem?

- Observation #1: SLOs are key – you need to count what counts
 - In many use cases, complete histories of measurements are not needed
 - Whatever was agreed to as part of an SLO
 - Capturing violations (and asserting their absence) is often sufficient (and more efficient to retain)
- Observation #2: Analogy between service and system failures
 - Failure to deliver precision is a failure
 - Precision availability is a form or availability



What Is PAM?

- Precision Availability Metrics express the availability of a service in accordance with the performance requirements reflected in a contract and expressed using Service Level Objectives (SLOs).
 - Example: a service with the requirement for not-to-exceed end-to-end latency
- SLEs, as defined in draft-ietf-teas-ietf-network-slices, are outside the scope because it is in the nature of SLEs that they define parts of the SLA that are difficult to measure.
- Performance requirements for various networking services can be expressed through a combination of SLOs. An SLO usually sets a threshold of one measurable metric that a service provider accepts as part of a service contract.

Composite Services and PAM

A composite service might include a set of connectivity constructs. An SLO might apply to all the constructs, or some constructs are assigned different sets of SLOs. For the purpose of PAM, each connectivity construct that composes the service can be monitored for its own SLO conformance as a sub-service. The composition of PAMs of these sub-services can be viewed as PAM of the composite service.

Applicability of PAM

Precision Availability Metrics (PAM) can be used:

- To determine the degree of compliance with which service levels are delivered relative to pre-defined SLOs.
- To provide service according to its SLO as part of accounting records; to account for the actual quality with which services were delivered and whether or not any SLO violations had occurred.
- To continuously monitor the quality with which the service is delivered.

Elements of PAM

- A PAM time unit, a.k.a. PAM interval, can be characterized as:
 - Violated Interval (VI) – an interval during which at least one of service level degraded below the pre-defined threshold
 - Violation-Free Interval (VFI) – all performance parameters are at or above their respective pre-defined optimal levels, and no defects have been detected
- Time interval: e.g., 1 second, or 1 msec
- Extensions possible, e.g., to differentiate “slight” and “severe” violations
 - Severely Violated Interval (SVI) – at least one of performance parameters degraded below the pre-defined critical threshold
- Based on these definitions, a set of basic metrics that count respective intervals is defined:
 - VI count, VFI count, and SVI count
- Violated packets can also be counted, but intervals are often more meaningful
 - Violations can occur in bursts: e.g., temporary overload conditions, route reconvergence
 - Differentiate “on rare occasions, sucks a lot” vs. “frequently, sucks just a little”
 - Compare Violated Intervals for transmissions

Derived PAM Metrics

- Based on basic PAM metrics a set of derived metrics is introduced for an VI:
 - Time since the last VI
 - Mean time between VIs
 - # Packets since the last VI
 - Mean # packets between VIs
- Analogous metrics introduced for SVI:
 - Time since the last SVI
 - Mean time between SVIs
 - # Packets since the last SVI
 - Mean # packets between SVIs

PAM extensions

- Account for lengthy disruptions, e.g.
 - Define significant duration threshold, e.g. ,10 time intervals
 - Extended unavailability metrics measure occurrence of consecutive VIs/SVIs beyond that threshold
- Complement with state model: service is deemed unavailable when the most recent intervals were all violated (or severely violated)
 - E.g., 10 consecutive SVIs constitute service unavailability state that begins at the start of the first SVI
 - E.g., 10 consecutive non-SVIs constitute service availability state that begins at the start of the first non-SVI
- Complement with additional derived metrics:
 - VI ratio – ratio of VIs to the total number of PAM intervals
 - SVI ratio – ratio of SVIs to the total number of PAM intervals

Discussion items

- Metrics: individual packets that violate SLO(s), e.g., counts of violations related to individual packets may be added in the future?

Future work (beyond this draft)

- YANG data model
- IPFIX Informational Elements
- Support for statistical SLOs, e.g., histogram and/or bucket
- Policies to define violated time unit, configure metrics
- Additional second-order metrics, e.g., “longest disruption of service time”

Next steps

- Welcome comments, questions (please include the IPPM WG ML).
- Always open to the cooperation.

Thank you!