Network Working Group G. Mirsky

Internet-Draft J. Halpern

Intended status: Standards Track Ericsson

Expires: 9 September 2023 X. Min

ZTE Corp.

A. Clemm

J. Strassner

Futurewei

J. Francois

Inria

8 March 2023

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

draft-ietf-ippm-pam-01

Abstract

This document defines a set of metrics for networking services with

performance requirements expressed as Service Level Objectives (SLO).

These metrics, referred to as Precision Availability Metrics (PAM),

are useful for defining and monitoring SLOs. For example, PAM can

be used by providers and/or users of a Network Slice service to

assess whether the service is provided in accordance with its defined SLOs.

Status of This Memo

This Internet-Draft is submitted in full conformance with the

provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering

Task Force (IETF). Note that other groups may also distribute

working documents as Internet-Drafts. The list of current Internet-

Drafts is at https://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months

and may be updated, replaced, or obsoleted by other documents at any

time. It is inappropriate to use Internet-Drafts as reference

material or to cite them other than as "work in progress."

This Internet-Draft will expire on 9 September 2023.

Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the

document authors. All rights reserved.

Mirsky, et al. Expires 9 September 2023 [Page 1]

Internet-Draft PAM for Multi-SLO March 2023

This document is subject to BCP 78 and the IETF Trust's Legal

Provisions Relating to IETF Documents (https://trustee.ietf.org/

license-info) in effect on the date of publication of this document.

Please review these documents carefully, as they describe your rights

and restrictions with respect to this document. Code Components

extracted from this document must include Revised BSD License text as

described in Section 4.e of the Trust Legal Provisions and are

provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction . . . . . . . . . . . . . . . . . . . . . . . . 2

2. Conventions and Terminology . . . . . . . . . . . . . . . . . 4

2.1. Terminology . . . . . . . . . . . . . . . . . . . . . . . 4

2.2. Acronyms . . . . . . . . . . . . . . . . . . . . . . . . 4

3. Precision Availability Metrics . . . . . . . . . . . . . . . 5

3.1. Introducing Violated Intervals . . . . . . . . . . . . . 5

3.2. Derived Precision Availability Metrics . . . . . . . . . 6

3.3. PAM Configuration Settings and Service Availability . . . 7

4. Statistical SLO . . . . . . . . . . . . . . . . . . . . . . . 8

5. Other PAM Benefits . . . . . . . . . . . . . . . . . . . . . 9

6. Extensions and Future Work . . . . . . . . . . . . . . . . . 9

7. IANA Considerations . . . . . . . . . . . . . . . . . . . . . 10

8. Security Considerations . . . . . . . . . . . . . . . . . . . 10

9. Acknowledgments . . . . . . . . . . . . . . . . . . . . . . . 11

10. References . . . . . . . . . . . . . . . . . . . . . . . . . 11

10.1. Informative References . . . . . . . . . . . . . . . . . 11

Contributors' Addresses . . . . . . . . . . . . . . . . . . . . . 12

Authors' Addresses . . . . . . . . . . . . . . . . . . . . . . . 12

1. Introduction

Network operators and users often need to assess the quality

with which network services are being provided and delivered. In

particular in cases where service level guarantees are given and

service level objectives (SLOs) are defined, it is essential to

provide a measure of the degree with which actual service levels that

are delivered comply with SLOs that were agreed, typically in a

contract or agreement. Examples of service levels include service

latency and packet loss. Examples of SLOs associated with

such service levels would be target values for the maximum packet

delay (one-way and/or round trip) or maximum packet loss ratio that

would be deemed acceptable.

An example of an SLO is one that characterizes the continued ability

of a particular set of nodes to communicate. Essentially, the

absence of what is called a defect, in other contexts. The SLO

would include the various time and measurement aspects that would be

Mirsky, et al. Expires 9 September 2023 [Page 2]

Internet-Draft PAM for Multi-SLO March 2023

interpreted as a defect or failure to communicate. It is important

to note that it is being defined as a state, and thus, it has

conditions that define entry into it and exit out of it. It is

expected that a Service Level Agreement (SLA) includes a defect-

related SLO, possibly in addition to other SLOs.

To express the perceived quality of delivered networking services

versus their SLOs, several metrics are needed to characterize the

quality of the service being provided. Of concern is not so much the

absolute service level (for example, actual latency experienced), but

whether the service is provided in accordance with the negotiated,

and eventually contracted, service levels. For instance, this may

include whether the packet delay that is experienced falls within an

acceptable range that has been contracted for the service. The

specific quality of service depends on the SLO that is in effect. A

non-conformance to an SLO might result in the degradation of the quality

of experience for gamers or even jeopardize the safety of a large

geographical area. However, as those applications represent clear

business opportunities, they demand dependable technical solutions.

The same service level may be deemed acceptable for one application,

while unacceptable for another, depending on the needs of the

application. Hence it is not sufficient to measure service

levels per se over time, but to assess the quality of the service

being contextually provided (e.g., with the applicable SLO(s) in mind).

However, at this point, there are no standard metrics in place that

can be used to account for the quality with which services are

delivered relative to their SLOs, and whether their SLOs are being

met at all times. Such metrics and the instrumentation to support

them are essential for various purposes, including monitoring (to

ensure that networking services are performing according to their

objectives) as well as accounting (to maintain a record of service

levels delivered, which is important for the monetization of such services as well

as for the triaging of problems).

The current state-of-the-art of metrics available today includes, for

example, interface metrics, useful to obtain statistical data on traffic volume

and behavior that can be observed at an interface [RFC2863] and

[RFC8343]. However, they are agnostic of actual service levels and

not specific to distinct flows. Flow records [RFC7011] and [RFC7012]

maintain statistics about flows, including flow volume and flow

duration, but again, contain very little information about end-to-end

service levels, let alone whether the service levels delivered

meet their targets, i.e., their associated SLOs.

This specification introduces a new set of metrics, Precision

Availability Metrics (PAM), aimed at capturing end-to-end service

levels for a flow, specifically the degree to which flows comply with

Mirsky, et al. Expires 9 September 2023 [Page 3]

Internet-Draft PAM for Multi-SLO March 2023

the SLOs that are in effect. PAM can be used to assess whether a

service is provided in compliance with its specified quality, i.e.,

in accordance with its defined SLOs. This information can be used in

multiple ways, for example, to optimize service delivery, take timely

counteractions in the event of service degradation, or account for

the quality of services being delivered.

Availability is discussed in Section 3.4 of [RFC7297]. In this

document, the term "availability" reflects that a service that is

characterized by its SLOs is considered unavailable whenever those

SLOs are violated, even if basic connectivity is still working.

"Precision" refers to the fact that services whose end-to-end service

levels are governed by SLOs, and which must therefore be precisely

delivered according to the associated quality and performance

requirements. It should be noted that precision refers to what is

being assessed, not the mechanism used to measure it; in other words,

it does not refer to the precision of the mechanism with which actual

service levels are measured. Furthermore, the precision, with

respect to the delivery of an SLO, only applies when the metric value

approaches the specified threshold levels in the SLO. The

specification and implementation of methods that provide for accurate

measurements is a separate topic independent of the definition of the

metrics in which the results of such measurements would be expressed.

Service Level Expectations (SLEs), as defined in Section 4.1 of

[I-D.ietf-teas-ietf-network-slices], are outside the scope of this

document, because it is in the nature of SLEs that they define parts

of the SLA that are not easily measured.

2. Conventions and Terminology

2.1. Terminology

In this document, SLA and SLO are used as defined in Section 4.1

[I-D.ietf-teas-ietf-network-slices].

2.2. Acronyms

PAM Precision Availability Metric

OAM Operations, Administration, and Maintenance

SLA Service Level Agreement

SLE Service Level Expectations

SLO Service Level Objective

Mirsky, et al. Expires 9 September 2023 [Page 4]

Internet-Draft PAM for Multi-SLO March 2023

VI Violated Interval

VIR Violated Interval Ratio

VPC Violated Packets Count

SVI Severely Violated Interval

SVIR Severely Violated Interval Ratio

SVPC Severely Violated Packets Count

VFI Violation-Free Interval

3. Precision Availability Metrics

3.1. Introducing Violated Intervals

When analyzing the availability metrics of a service flow between two

nodes, we need to select a time interval as the unit of PAM. In

[ITU.G.826], a time interval of one second is used. That is

reasonable, but some services may require different granularity. For

that reason, the time interval in PAM is viewed as a variable

parameter though constant for a particular measurement session.

Further, for the purpose of PAM, each time interval, e.g., second or

decamillisecond, is classified either as Violated Interval (VI),

Severely Violated Interval (SVI), or Violation-Free Interval (VFI).

These are defined as follows:

\* VI is a time interval during which at least one of the performance

parameters degraded below its pre-defined optimal level threshold.

\* SVI is a time interval during which at least one the performance

parameters degraded below its pre-defined critical threshold.

\* Consequently, VFI is a time interval during which all performance

objectives are at or better than their respective pre-defined

optimal levels.

Mechanisms of setting levels of threshold of an SLO are outside the

scope of this document.

From these definitions, a set of basic metrics can be defined that

count the numbers of time intervals that fall into each category:

\* VI count.

\* SVI count.

Mirsky, et al. Expires 9 September 2023 [Page 5]

Internet-Draft PAM for Multi-SLO March 2023

\* VFI count.

These count metrics are essential in calculating respective ratios

(see Section 3.2) that can be used to assess the instability of the

service.

Beyond accounting for violated intervals, it can sometimes be

beneficial also to maintain counts of packets for which a performance

threshold is violated. For example, this allows to distinguish

between cases in which violated intervals are caused by isolated

violation occurrences (such as, a sporadic issue that may be caused

by a temporary spike in a queue depth along the packet's path) or by

broad violations across multiple packets (such as a problem with slow

route reconvergence across the network or more foundational issues

such as insufficient network resources). Maintaining such counts and

comparing them with the overall amount of traffic also facilitates

assessing compliance with statistical SLOs (see Section 4). For

these reasons, the following additional metrics are defined:

\* VPC: Violated packets count

\* SVPC: Severely violated packets count

3.2. Derived Precision Availability Metrics

A set of metrics can be created based on PAM introduced in Section 3.

In this document, these metrics are referred to as derived PAM. Some

of these metrics are modeled after Mean Time Between Failure (MTBF)

metrics - a "failure" in this context referring to a failure to

deliver a packet according to its SLO.

\* Time since the last violated interval (e.g., since last violated

ms, since last violated second). (This parameter is suitable for

monitoring the current compliance status of the service, e.g., for

trending analysis.)

\* Number of packets since the last violated packet. (This parameter

is suitable for the monitoring of the current compliance status of

the service.)

\* Mean time between VIs (e.g., between violated milliseconds,

violated seconds) is the arithmetic mean of time between

consecutive VIs.

\* Mean packets between VIs is the arithmetic mean of the number of

SLO-compliant packets between consecutive VIs. (Another variation

of "MTBF" in a service setting.)

Mirsky, et al. Expires 9 September 2023 [Page 6]

Internet-Draft PAM for Multi-SLO March 2023

An analogous set of metrics can be produced for SVI:

\* Time since the last SVI (e.g., since last violated ms, since last

violated second). (This parameter is suitable for the monitoring

of the current compliance status of the service.)

\* Number of packets since the last severely violated packet. (This

parameter is suitable for the monitoring of the current compliance

status of the service.)

\* Mean time between SVIs (e.g., between severely violated

milliseconds, severely violated seconds) is the arithmetic mean of

time between consecutive SVIs.

\* Mean packets between SVIs is the arithmetic mean of the number of

SLO-compliant packets between consecutive SVIs. (Another

variation of "MTBF" in a service setting.)

To indicate a historic degree of precision availability, additional

derived PAMs can be defined as follows:

\* Violated interval ratio (VIR) is the ratio of the combined numbers

of VIs and SVIs to the total number of time unit intervals in a

time of the availability periods during a fixed measurement

interval.

\* Severely violated interval ratio (SVIR) - is the ratio of SVIs to

the total number of time unit intervals in a time of the

availability periods during a fixed measurement interval.

3.3. PAM Configuration Settings and Service Availability

It might be useful for a network operator to determine the current

condition of the service for which Precision Availability Metrics are

maintained. To facilitate this, it is conceivable to complement PAM

with a state model. Such a state model can be used to indicate

whether a service is currently considered available or unavailable

depending on the network's recent ability to provide service without

incurring intervals during which violations occur. It is conceivable

to define such a state model in which transitions occur per some

predefined PAM settings.

While the definition of a service state model is outside the scope of

this draft, the following section provides some considerations for

how such a state model and accompanying configuration settings could

be defined.

Mirsky, et al. Expires 9 September 2023 [Page 7]

Internet-Draft PAM for Multi-SLO March 2023

For example, a state model could be defined by a Finite State Machine

featuring two states, "available" and "unavailable". The initial

state could be "available". A service could subsequently be deemed

as "unavailable" based on the number of successive interval

violations that have been (recently) experienced. To return to a state

of "available", a number of intervals without violations would need

to be observed.

The number of successive intervals with violations, as well as the

number of successive intervals that are free of violations, required

for a state to transition to another state is defined by a

configuration setting. Specifically, the following configuration

parameters could be defined:

\* Unavailability threshold: The number of successive intervals

during which a violation occurs to transition to an unavailable

state.

\* Availability threshold: The number of successive intervals during

which no violations must occur to allow transition to an available

state from a previously unavailable state.

Additional configuration parameters could be defined to account for

the severity of violations. Likewise, it is conceivable to define

configuration settings that also take VIR and SVIR into account.

4. Statistical SLO

It should be noted that certain SLAs may be statistical, requiring

the service levels of packets in a flow to adhere to specific

distributions. For example, an SLA might state that any given SLO

applies to at least a certain percentage of packets, allowing for a

certain level of, for example, packet loss and/or exceeding packet

delay threshold to take place. Each such event, in that case, does

not necessarily constitute an SLO violation. However, it is still

useful to maintain those statistics, as the number of out-of-SLO

packets still matters when looked at in proportion to the total

number of packets.

Along that vein, an SLA might establish an SLO of, say, end-to-end

latency to not exceed 20 ms for 99% of packets, to not exceed 25ms

for 99.999% of packets, and to never exceed 30ms for any packet. In

that case, any individual packet with latency greater than 20 ms

latency and lower than 30 ms cannot be considered an SLO violation in

itself, but compliance with the SLO may need to be assessed after the

fact.

Mirsky, et al. Expires 9 September 2023 [Page 8]

Internet-Draft PAM for Multi-SLO March 2023

To support statistical SLOs more directly requires additional

metrics, such as metrics that represent histograms for service level

parameters with buckets corresponding to individual service level

objectives. For the example just given, a histogram for a given flow

could be maintained with three buckets: one containing the count of

packets within 20ms, a second with a count of packets between 20 and

25ms (or simply all within 25ms), a third with a count of packets

between 25 and 30ms (or merely all packets within 30ms, and a fourth

with a count of anything beyond (or simply a total count). Of

course, the number of buckets and the boundaries between those

buckets should correspond to the needs of the SLA associated with the

application, i.e., to the specific guarantees and SLOs that were

provided. The definition of histogram metrics is for further study

(see Section 6).

5. Other PAM Benefits

PAM provides several benefits with other, more conventional

performance metrics. Without PAM, it would be possible to conduct

ongoing measurements of service levels and maintain a time-series of

service level records, then assess compliance with specific SLOs

after the fact. However, doing so would require the collection of

vast amounts of data that would need to be generated, exported,

transmitted, collected, and stored. In addition, extensive

postprocessing would be required to compare that data against SLOs

and analyze its compliance. Being able to perform these tasks at

scale and in real-time would present significant additional

challenges.

Adding PAM allows for a more compact expression of service level

compliance. In that sense, PAM does not simply represent raw data

but expresses actionable information. In conjunction with proper

instrumentation, PAM can thus help avoid expensive postprocessing.

6. Extensions and Future Work

The following is a list of items that are outside the scope of this

specification, but which will be useful extensions and opportunities

for future work:

\* A YANG data model will allow PAM to be incorporated into

monitoring applications based on the YANG/NETCONF/RESTCONF

framework. In addition, a YANG data model will enable the

configuration of PAM-related settings.

Mirsky, et al. Expires 9 September 2023 [Page 9]

Internet-Draft PAM for Multi-SLO March 2023

\* A set of IPFIX Information Elements will allow Precision

Availability Metrics to be associated with flow records and

exported as part of flow data, for example for processing by

accounting applications that assess compliance of delivered

services with quality guarantees.

\* Additional second-order metrics, such as "longest disruption of

service time" (measuring consecutive time units with SVIs), can be

defined and would be deemed useful by some users. At the same

time, such metrics can be computed in a straightforward manner and

will in many cases be application-specific. For this reason,

further such metrics are omitted here in order to not overburden

this specification.

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

Instrumentation for metrics that are used to assess compliance with

SLOs constitute an attractive target for an attacker. By interfering

with the maintenance of such metrics, services could be falsely

identified as complying (when they are not) or vice-versa (i.e.,

flagged as being non-compliant when indeed they are). While this

document does not specify how networks should be instrumented to

maintain the identified metrics, such instrumentation needs to be

adequately secured to ensure accurate measurements and prohibit

tampering with metrics being kept.

Where metrics are being defined relative to an SLO, the configuration

of those SLOs needs to be adequately secured. Likewise, where SLOs

can be adjusted, the correlation between any metric instance and a

particular SLO must be unambiguous. The same service levels that

constitute SLO violations for one flow that should be maintained as

part of the "violated time units" and related metrics, may be

compliant for another. In cases when it is impossible

to tie together SLOs and PAM, it will be preferable to

merely maintain statistics about service levels delivered (for

example, overall histograms of end-to-end latency) without assessing

which constitutes violations.

By the same token, where the definition of what constitutes a

"severe" or a "significant" violation depends on configuration

settings or context. The configuration of such settings or context

needs to be specially secured. Also, the configuration must be bound

to the metrics being maintained. Thus, it will be clear which

configuration setting was in effect when those metrics were being

Mirsky, et al. Expires 9 September 2023 [Page 10]

Internet-Draft PAM for Multi-SLO March 2023

assessed. An attacker that can tamper with such configuration

settings will render the corresponding metrics useless (in the best

case) or misleading (in the worst case).

9. Acknowledgments

TBA

10. References

10.1. Informative References

[I-D.ietf-teas-ietf-network-slices]

Farrel, A., Drake, J., Rokui, R., Homma, S., Makhijani,

K., Contreras, L. M., and J. Tantsura, "A Framework for

IETF Network Slices", Work in Progress, Internet-Draft,

draft-ietf-teas-ietf-network-slices-19, 21 January 2023,

<https://datatracker.ietf.org/doc/html/draft-ietf-teas-

ietf-network-slices-19>.

[ITU.G.826]

ITU-T, "End-to-end error performance parameters and

objectives for international, constant bit-rate digital

paths and connections", ITU-T G.826, December 2002.

[RFC2863] McCloghrie, K. and F. Kastenholz, "The Interfaces Group

MIB", RFC 2863, DOI 10.17487/RFC2863, June 2000,

<https://www.rfc-editor.org/info/rfc2863>.

[RFC7011] Claise, B., Ed., Trammell, B., Ed., and P. Aitken,

"Specification of the IP Flow Information Export (IPFIX)

Protocol for the Exchange of Flow Information", STD 77,

RFC 7011, DOI 10.17487/RFC7011, September 2013,

<https://www.rfc-editor.org/info/rfc7011>.

[RFC7012] Claise, B., Ed. and B. Trammell, Ed., "Information Model

for IP Flow Information Export (IPFIX)", RFC 7012,

DOI 10.17487/RFC7012, September 2013,

<https://www.rfc-editor.org/info/rfc7012>.

[RFC7297] Boucadair, M., Jacquenet, C., and N. Wang, "IP

Connectivity Provisioning Profile (CPP)", RFC 7297,

DOI 10.17487/RFC7297, July 2014,

<https://www.rfc-editor.org/info/rfc7297>.

[RFC8343] Bjorklund, M., "A YANG Data Model for Interface

Management", RFC 8343, DOI 10.17487/RFC8343, March 2018,

<https://www.rfc-editor.org/info/rfc8343>.

Mirsky, et al. Expires 9 September 2023 [Page 11]

Internet-Draft PAM for Multi-SLO March 2023

Contributors' Addresses

Liuyan Han

China Mobile

32 XuanWuMenXi Street

Beijing

100053

China

Email: hanliuyan@chinamobile.com

Mohamed Boucadair

Orange

35000 Rennes

France

Email: mohamed.boucadair@orange.com

Adrian Farrel

Old Dog Consulting

United Kingdom

Email: adrian@olddog.co.uk

Authors' Addresses

Greg Mirsky

Ericsson

Email: gregimirsky@gmail.com

Joel Halpern

Ericsson

Email: joel.halpern@ericsson.com

Xiao Min

ZTE Corp.

Email: xiao.min2@zte.com.cn

Alexander Clemm

Futurewei

2330 Central Expressway

Santa Clara, CA 95050

United States of America

Email: ludwig@clemm.org

Mirsky, et al. Expires 9 September 2023 [Page 12]

Internet-Draft PAM for Multi-SLO March 2023

John Strassner

Futurewei

2330 Central Expressway

Santa Clara, CA 95050

United States of America

Email: strazpdj@gmail.com

Jerome Francois

Inria

615 Rue du Jardin Botanique

54600 Villers-les-Nancy

France

Email: jerome.francois@inria.fr

Mirsky, et al. Expires 9 September 2023 [Page 13]