

Internet Engineering Task Force (IETF)
Request for Comments: 9544
Category: Informational
ISSN: 2070-1721

G. Mirsky
J. Halpern
Ericsson
X. Min
ZTE Corp.
A. Clemm

J. Strassner
Futurewei
J. Francois
Inria and University of Luxembourg
March 2024

Precision Availability Metrics (PAMs) for Services Governed by Service Level Objectives (SLOs)

Abstract

This document defines a set of metrics for networking services with performance requirements expressed as Service Level Objectives (SLOs). These metrics, referred to as "Precision Availability Metrics (PAMs)", are useful for defining and monitoring SLOs. For example, PAMs can be used by providers and/or customers of an RFC 9543 Network Slice Service to assess whether the service is provided in compliance with its defined SLOs.

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 candidates for any level of Internet Standard; see 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/rfc9544>.

Copyright Notice

Copyright (c) 2024 IETF Trust and the persons identified as the document authors. All rights reserved.

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. Conventions

2.1.	Terminology
2.2.	Acronyms
3.	Precision Availability Metrics
3.1.	Introducing Violated Intervals
3.2.	Derived Precision Availability Metrics
3.3.	PAM Configuration Settings and Service Availability
4.	Statistical SLO
5.	Other Expected PAM Benefits
6.	Extensions and Future Work
7.	IANA Considerations
8.	Security Considerations
9.	Informative References
	Acknowledgments
	Contributors
	Authors' Addresses

1. Introduction

Service providers and users often need to assess the quality with which network services are being delivered. In particular, in cases where service-level guarantees are documented (including their companion metrology) as part of a contract established between the customer and the service provider, and Service Level Objectives (SLOs) are defined, it is essential to provide means to verify that what has been delivered complies with what has been possibly negotiated and (contractually) defined between the customer and the service provider. Examples of SLOs 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.

More generally, SLOs can be used to characterize the ability of a particular set of nodes to communicate according to certain measurable expectations. Those expectations can include but are not limited to aspects such as latency, delay variation, loss, capacity/throughput, ordering, and fragmentation. Whatever SLO parameters are chosen and whichever way service-level parameters are being measured, Precision Availability Metrics indicate whether or not a given service has been available according to expectations at all times.

Several metrics (often documented in the IANA "Performance Metrics" registry [IANA-PM-Registry] according to [RFC8911] and [RFC8912]) can be used to characterize the service quality, expressing the perceived quality of delivered networking services versus their SLOs. Of concern is not so much the absolute service level (for example, actual latency experienced) but whether the service is provided in compliance with the negotiated and eventually contracted service levels. For instance, this may include whether the experienced packet delay falls within an acceptable range that has been contracted for the service. The specific quality of service depends on the SLO or a set thereof for a given service that is in effect. Non-compliance 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.

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; the quality of the service being contextually provided (e.g., with the applicable SLO in mind) must be also assessed. However, at this point, there are no standard metrics that can be used to account for the quality with which services are delivered relative to their SLOs or to determine 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 include, for example, interface metrics that can be used to obtain statistical data on traffic volume and behavior that can be observed at an interface [RFC2863] [RFC8343]. However, they are agnostic of actual service levels and not specific to distinct flows. Flow records [RFC7011] [RFC7012] maintain statistics about flows, including flow volume and flow duration, but again, they contain very little information about service levels, let alone whether the service levels delivered meet their respective targets, i.e., their associated SLOs.

This specification introduces a new set of metrics, Precision Availability Metrics (PAMs), aimed at capturing service levels for a flow, specifically the degree to which the flow complies with the SLOs that are in effect. PAMs can be used to assess whether a service is provided in compliance 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 services whose service levels are governed by SLOs and must be delivered precisely 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, particularly applies when a metric value approaches the specified threshold levels in the SLO.

The specification and implementation of methods that provide for accurate measurements are separate topics independent of the definition of the metrics in which the results of such measurements would be expressed. Likewise, Service Level Expectations (SLEs), as defined in Section 5.1 of [RFC9543], are outside the scope of this document.

2. Conventions

2.1. Terminology

In this document, SLA and SLO are used as defined in [RFC3198]. The reader may refer to Section 5.1 of [RFC9543] for an applicability example of these concepts in the context of RFC 9543 Network Slice Services.

2.2. Acronyms

IPFIX	IP Flow Information Export
PAM	Precision Availability Metric
SLA	Service Level Agreement
SLE	Service Level Expectation
SLO	Service Level Objective
SVI	Severely Violated Interval

SVIR Severely Violated Interval Ratio

SVPC Severely Violated Packets Count

VFI Violation-Free Interval

VI Violated Interval

VIR Violated Interval Ratio

VPC Violated Packets Count

3. Precision Availability Metrics

3.1. Introducing Violated Intervals

When analyzing the availability metrics of a service between two measurement points, a time interval as the unit of PAMs needs to be selected. In [ITU.G.826], a time interval of one second is used. That is reasonable, but some services may require different granularity (e.g., decamillisecond). For that reason, the time interval in PAMs is viewed as a variable parameter, though constant for a particular measurement session. Furthermore, for the purpose of PAMs, each time interval is classified as either 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 configurable optimal threshold.
- * SVI is a time interval during which at least one of the performance parameters degraded below its configurable critical threshold.
- * Consequently, VFI is a time interval during which all performance parameters are at or better than their respective pre-defined optimal levels.

The monitoring of performance parameters to determine the quality of an interval is performed between the elements of the network that are identified in the SLO corresponding to the performance parameter. Mechanisms for setting levels of a threshold of an SLO are outside the scope of this document.

From the definitions above, a set of basic metrics can be defined that count the number of time intervals that fall into each category:

- * VI count
- * SVI count
- * VFI count

These count metrics are essential in calculating respective ratios (see Section 3.2) that can be used to assess the instability of a service.

Beyond accounting for violated intervals, it is sometimes beneficial to maintain counts of packets for which a performance threshold is violated. For example, this allows for distinguishing 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 convergence 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 facilitate 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 PAMs as introduced in this document. In this document, these metrics are referred to as "derived PAMs". Some of these metrics are modeled after Mean Time Between Failure (MTBF) metrics; a "failure" in this context refers to a failure to deliver a service according to its SLO.

- * Time since the last violated interval (e.g., since last violated ms or 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 or between violated seconds). This parameter is the arithmetic mean of time between consecutive VIs.
- * Mean packets between VIs. This parameter is the arithmetic mean of the number of SLO-compliant packets between consecutive VIs. It is another variation of MTBF in a service setting.

An analogous set of metrics can be produced for SVI:

- * Time since the last SVI (e.g., since last violated ms or 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 or between severely violated seconds). This parameter is the arithmetic mean of time between consecutive SVIs.
- * Mean packets between SVIs. This parameter is the arithmetic mean of the number of SLO-compliant packets between consecutive SVIs. It is 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 summed numbers of VIs and SVIs to the total number of time unit intervals in a time of the availability periods during a fixed measurement session.
- * 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 session.

3.3. PAM Configuration Settings and Service Availability

It might be useful for a service provider to determine the current

condition of the service for which PAMs are maintained. To facilitate this, it is conceivable to complement PAMs with a state model. Such a state model can be used to indicate whether a service is currently considered as 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 document, this section provides some considerations for how such a state model and accompanying configuration settings could be defined.

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 experienced up to the particular observation time moment. 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 are 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 a multi-tiered SLO of, say, end-to-end latency (from the lowest to highest tier) as follows:

- * not to exceed 30 ms for any packet;
- * not to exceed 25 ms for 99.999% of packets; and
- * not to exceed 20 ms for 99% of packets.

In that case, any individual packet with a 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.

To support statistical SLOs more directly requires additional metrics, for example, metrics that represent histograms for service-level parameters with buckets corresponding to individual SLOs. Although the definition of histogram metrics is outside the scope of this document and could be considered for future work (see Section 6), for the example just given, a histogram for a particular flow could be maintained with four buckets: one containing the count of packets within 20 ms, a second with a count of packets between 20 and 25 ms (or simply all within 25 ms), a third with a count of packets between 25 and 30 ms (or merely all packets within 30 ms), 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.

5. Other Expected PAM Benefits

PAMs provide several benefits with other, more conventional performance metrics. Without PAMs, it would be possible to conduct ongoing measurements of service levels, maintain a time series of service-level records, and 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 post-processing 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 PAMs allows for a more compact expression of service-level compliance. In that sense, PAMs do not simply represent raw data but expresses actionable information. In conjunction with proper instrumentation, PAMs can thus help avoid expensive post-processing.

6. Extensions and Future Work

The following is a list of items that are outside the scope of this specification but will be useful extensions and opportunities for future work:

- * A YANG data model will allow PAMs to be incorporated into monitoring applications based on the YANG, NETCONF, and RESTCONF frameworks. In addition, a YANG data model will enable the configuration and retrieval of PAM-related settings.
- * A set of IPFIX Information Elements will allow PAMs 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 be application specific in many cases. For this reason, such metrics are omitted here in order to not overburden this specification.
- * Metrics can be defined to represent histograms for service-level parameters with buckets corresponding to individual SLOs.

7. IANA Considerations

This document has no IANA actions.

8. Security Considerations

Instrumentation for metrics that are used to assess compliance with SLOs constitutes 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 and should be maintained as part of the "violated time units" and related metrics may be compliant for another flow. In cases when it is impossible to tie together SLOs and PAMs, it is preferable to merely maintain statistics about service levels delivered (for example, overall histograms of end-to-end latency) without assessing which constitute violations.

By the same token, 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 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. Informative References

[IANA-PM-Registry]

IANA, "Performance Metrics",
<<https://www.iana.org/assignments/performance-metrics>>.

[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>>.

[RFC3198]

Westerinen, A., Schnizlein, J., Strassner, J., Scherling, M., Quinn, B., Herzog, S., Huynh, A., Carlson, M., Perry, J., and S. Waldbusser, "Terminology for Policy-Based Management", RFC 3198, DOI 10.17487/RFC3198, November 2001, <<https://www.rfc-editor.org/info/rfc3198>>.

[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>>.
- [RFC8911] Bagnulo, M., Claise, B., Eardley, P., Morton, A., and A. Akhter, "Registry for Performance Metrics", RFC 8911, DOI 10.17487/RFC8911, November 2021, <<https://www.rfc-editor.org/info/rfc8911>>.
- [RFC8912] Morton, A., Bagnulo, M., Eardley, P., and K. D'Souza, "Initial Performance Metrics Registry Entries", RFC 8912, DOI 10.17487/RFC8912, November 2021, <<https://www.rfc-editor.org/info/rfc8912>>.
- [RFC9543] Farrel, A., Ed., Drake, J., Ed., Rokui, R., Homma, S., Makhijani, K., Contreras, L., and J. Tantsura, "A Framework for Network Slices in Networks Built from IETF Technologies", RFC 9543, DOI 10.17487/RFC9543, March 2024, <<https://www.rfc-editor.org/info/rfc9543>>.

Acknowledgments

The authors greatly appreciate review and comments by Bjørn Ivar Teigen and Christian Jacquenet.

Contributors

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
Email: ludwig@clemm.org

John Strassner
Futurewei
2330 Central Expressway
Santa Clara, CA 95050
United States of America
Email: strazpdj@gmail.com

Jerome Francois
Inria and University of Luxembourg
615 Rue du Jardin Botanique
54600 Villers-les-Nancy
France
Email: jerome.francois@inria.fr