

Network Working Group
Internet-Draft
Intended status: BCP
Expires: September 5, 2009

A. Clark
Telchemy Incorporated
March 9, 2009

Framework for Performance Metric Development
draft-ietf-pmol-metrics-framework-02

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/1id-abstracts.txt>.

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>.

This Internet-Draft will expire on September 9, 2009.

Copyright Notice

Copyright (c) 2009 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 in effect on the date of publication of this document (<http://trustee.ietf.org/license-info>). Please review these documents carefully, as they describe your rights and restrictions with respect to this document.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format

it for publication as an RFC or to translate it into languages other than English.

Clark

[Page 1]

Metrics Framework

9 March 2009

Abstract

This memo describes a framework and guidelines for the development of performance metrics that are beyond the scope of existing working group charters in the IETF. In this version, the memo refers to a Performance Metrics Entity, or PM Entity, which may in future be a working group or directorate or a combination of these two.

Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC 2119](#) [[RFC2119](#)].

Table of Contents

1.	Introduction	2
1.1.	Background and Motivation	3
1.2.	Organization of this memo	3
2.	Purpose and Scope	4
3.	Metrics Development	4
3.1.	Audience for Metrics	4
3.2.	Definitions of a Metric	4
3.3.	Computed Metrics	5
3.3.1	Composed Metrics	5
3.3.2	Index	5
3.4.	Metric Specification	6
3.4.1.	Outline	6
3.4.2.	Normative parts of metric definition	6
3.4.3.	Informative parts of metric definition	7
3.4.4.	Metric Definition Template	8
3.4.5.	Examples	8
3.5.	Dependencies	9
3.5.1.	Timing accuracy	9
3.5.2.	Dependencies of metric definitions on related events or metrics	9
3.5.3.	Relationship between application performance and lower layer metrics	10
4.	Performance Metric Development Process	10
4.1.	New Proposals for Metrics	11
4.2.	Reviewing Metrics	11
4.3.	Proposal Approval	11
4.4.	PM Entity Interaction with other WGs	11
4.5.	Standards Track Performance Metrics	11

5.	IANA Considerations	12
6.	Security Considerations	12
7.	Acknowledgements	12
8.	References	12
8.1.	Normative References	12
8.2.	Informative References	12
	Author's Address	13
	Intellectual Property and Copyright Statements	13

1. Introduction

Many applications are distributed in nature, and their performance may be impacted by a IP impairments, server capacity, congestion and other factors. It is important to measure the performance of applications and services to ensure that quality objectives are being met and to support problem diagnosis. Standardized metrics help to ensure that performance measurement is implemented consistently and to facilitate interpretation and comparison.

There are at least three phases in the development of performance standards. They are:

- 1. Definition of a Performance Metric and its units of measure
- 2. Specification of a Method of Measurement
- 3. Specification of the Reporting Format

During the development of metrics it is often useful to define performance objectives and expected value ranges however this is not defined as part of the metric specification.

This memo refers to a Performance Metrics Entity, or PM Entity, which may in future be a working group or directorate or a combination of these two.

1.1. Background and Motivation

Although the IETF has two active Working Groups dedicated to the development of performance metrics, they each have strict limitations in their charters:

- The Benchmarking Methodology WG has addressed a range of networking technologies and protocols in their long history (such as IEEE 802.3, ATM, Frame Relay, and Routing Protocols), but the charter strictly limits their performance characterizations to the laboratory environment.

- The IP Performance Metrics WG has the mandate to develop metrics applicable to live IP networks, but it is specifically prohibited from developing metrics that characterize traffic (such as a VoIP stream).

A BOF held at IETF-69 introduced the IETF community to the possibility of a generalized activity to define standardized performance metrics. The existence of a growing list of Internet-Drafts on performance metrics (with community interest in development, but in un-chartered areas) illustrates the need for additional performance work. The majority of people present at the BOF supported the proposition that IETF should be working in these areas, and no one objected to any of the proposals.

Clark

[Page 3]

Metrics Framework

9 March 2009

The IETF does have current and completed activities related to the reporting of application performance metrics (e.g. RAQMOM) and is also actively involved in the development of reliable transport protocols which would affect the relationship between IP performance and application performance.

Thus there is a gap in the currently chartered coverage of IETF WGs: development of performance metrics for non-IP-layer protocols that can be used to characterize performance on live networks.

1.2. Organization of this memo

This memo is divided in two major sections beyond the Purpose and Scope section. The first is a definition and description of a performance metric and its key aspects. The second defines a process to develop these metrics that is applicable to the IETF environment.

2. Purpose and Scope

The purpose of this memo is to define a framework and a process for developing performance metrics for IP-based applications that operate over reliable or datagram transport protocols, and that can be used to characterize traffic on live networks and services.

The scope of this memo includes the support of metric definition for any protocol developed by the IETF, however this memo is not intended to supercede existing working methods within WGs that have existing chartered work in this area.

This process is not intended to govern performance metric development in existing IETF WG that are focused on metrics development, such as IPPM and BMWG. However, the framework and guidelines may be useful in these activities, and MAY be applied where appropriate.

3. Metrics Development

This section provides key definitions and qualifications of performance metrics.

3.1. Audience for Metrics

Metrics are intended for use in measuring the performance of an application, network or service. A key first step in metric definition is to identify what metrics are needed by the "user" in order to properly maintain service quality and to identify and quantify problems, i.e. to consider the audience for the metrics.

3.2. Definitions of a Metric

A metric is a measure of an observable behavior of an application, protocol or other system. The definition of a metric often assumes some implicit or explicit underlying statistical process, and a metric is an estimate of a parameter of this process. If the assumed statistical process closely models the behavior of the system then

Clark

[Page 4]

Metrics Framework

9 March 2009

the metric is "better" in the sense that it more accurately characterizes the state or behavior of the system.

A metric should serve some defined purpose. This may include the measurement of capacity, quantifying how bad some problem is, measurement of service level, problem diagnosis or location and other such uses. A metric may also be an input to some other process, for example the computation of a composite metric or a model or simulation of a system. Tests of the "usefulness" of a metric include:

- (i) the degree to which its absence would cause significant loss of information on the behavior or state of the application or system being measured
- (ii) the correlation between the metric and the QoS [[G1000](#)] / experience delivered to the user (person or other application)
- (iii) the degree to which the metric is able to support the identification and location of problems affecting service quality.

For example, consider a distributed application operating over a network connection that is subject to packet loss. A Packet Loss Rate (PLR) metric is defined as the mean packet loss rate over some time period. If the application performs poorly over network connections with high packet loss rate and always performs well when the packet loss rate is zero then the PLR metric is useful to some degree. Some applications are sensitive to short periods of high loss (bursty loss) and are relatively insensitive to isolated packet

loss events; for this type of application there would be very weak correlation between PLR and application performance. A "better" metric would consider both the packet loss rate and the distribution of loss events. If application performance is degraded when the PLR exceeds some rate then a useful metric may be a measure of the duration and frequency of periods during which the PLR exceeds that rate.

3.3. Computed Metrics

3.3.1 Composed Metrics

Some metrics may not be measured directly, but may be composed from metrics that have been measured. Usually the contribution metrics have a limited scope in time or space, and they can be combined to estimate the performance of some larger entity. Some examples of composed metrics and composed metric definitions are:

Spatial Composition is defined as the composition of metrics of the same type with differing spatial domains [Ref ?]. For spatially composed metrics to be meaningful, the spatial domains should be non-overlapping and contiguous, and the composition operation should be mathematically appropriate for the type of metric.

Clark

[Page 5]

Metrics Framework

9 March 2009

Temporal Composition is defined as the composition of sets of metrics of the same type with differing time spans [Ref ?]. For temporally composed metrics to be meaningful, the time spans should be non-overlapping and contiguous, and the composition operation should be mathematically appropriate for the type of metric.

Temporal Aggregation is a summarization of metrics into a smaller number of metrics that relate to the total time span covered by the original metrics. An example would be to compute the minimum, maximum and average values of a series of time sampled values of a metric.

3.3.2 Index (from compagg)

An Index is a metric for which the output value range has been selected for convenience or clarity, and the behavior of which is selected to support ease of understanding (e.g. G.107 R Factor). The deterministic function for an index is often developed after the index range and behavior have been determined.

3.4. Metric Specification

3.4.1. Outline

A metric definition MUST have a normative part that defines what the

metric is and how it is measured or computed and SHOULD have an informative part that describes the metric and its application.

3.4.2. Normative parts of metric definition

The normative part of a metric definition MUST define at least the following:

(i) Metric Name

Metric names MUST be unique within the set of metrics being defined and MAY be descriptive.

(ii) Metric Description

The description MUST explain what the metric is, what is being measured and how this relates to the performance of the system being measured.

(iii) Measurement Method

This MUST define what is being measured, estimated or computed and the specific algorithm to be used. Terms such as "average" should be qualified (e.g. running average or average over some interval). Exception cases SHOULD also be defined with the appropriate handling method. For example, there are a number of commonly used metrics related to packet loss; these often don't define the criteria by which a packet is determined to be lost (vs very delayed) or how duplicate packets are handled. For example, if the average packet loss rate during a time interval is reported, and a packet's arrival is delayed from one interval to the next then was it "lost" during the interval during which it should have arrived or should it be counted as received?

Clark

[Page 6]

Metrics Framework

9 March 2009

(iv) Units of measurement

The units of measurement MUST be clearly stated.

(v) Measurement timing

The acceptable range of timing intervals or sampling intervals for a measurement and the timing accuracy required for such intervals MUST be specified. Short sampling intervals or frequent samples provide a rich source of information that can help to assess application performance but may lead to excessive measurement data. Long measurement or sampling intervals reduce the amount of reported and collected data however may be insufficient to truly understand potentially time varying application performance or service quality.

(vi) Measurement Point

If the measurement is specific to a measurement point this SHOULD be defined.

3.4.3. Informative parts of metric definition

The informative part of a metric specification is intended to support the implementation and use of the metric. This part SHOULD provide the following data:

(i) Implementation

The implementation description MAY be in the form of text, algorithm or example software. The objective of this part of the metric definition is to assist implementers to achieve a consistent result.

(ii) Verification

The metric definition SHOULD provide guidance on verification testing. This may be in the form of test vectors, a formal verification test method or informal advice.

(iii) Use and Applications

The Use and Applications description is intended to assist the "user" to understand how, when and where the metric can be applied, and what significance the value range for the metric may have. This MAY include a definition of the "typical" and "abnormal" range of the metric, if this was not apparent from the nature of the metric. For example:

(a) it is fairly intuitive that a lower packet loss rate would equate to better performance however the user may not know the significance of some given packet loss rate,

(b) the speech level of a telephone signal is commonly expressed in dBm0. If the user is presented with:

Speech level = -7 dBm0

Clark

[Page 7]

Metrics Framework

9 March 2009

this is not intuitively understandable, unless the user is a telephony expert. If the metric definition explains that the typical range is -18 to -28 dBm0, a value higher than -18 means the signal may be too high (loud) and less than -28 means that the signal may be too low (quiet), it is much easier to interpret the metric.

(iv) Reporting Model

The Reporting Model definition is intended to make any relationship between the metric and the reporting model clear. There are often implied relationships between the method of reporting metrics and the metric itself, however these are often not made apparent to the

implementor. For example, if the metric is a short term running average packet delay variation (e.g. PPDV as defined in [RFC3550](#)) that is reported at intervals of 6-10 seconds the resulting measurement may have limited accuracy if packet delay variation is non-stationary.

3.4.4. Metric Definition Template

Normative

Metric Name
Metric Description
Measurement Method
Units of measurement
Measurement Timing

Informative

Implementation Guidelines
Verification
Use and Applications
Reporting Model

3.4.5. Examples

Example definition

Metric Name: BurstPacketLossFrequency

Metric Description: A burst of packet loss is defined as a longest period starting and ending with lost packets during which no more than Gmin consecutive packets are received. The BurstPacketLossFrequency is defined as the number of bursts of packet loss occurring during a specified time interval (e.g. per minute, per hour, per day). If Gmin is set to 0 then a burst of packet loss would comprise only consecutive lost packets, whereas a Gmin of 16 would define bursts as periods of both lost and received packets (sparse bursts) having a loss rate of greater than 5.9%.

Measurement Method: Bursts may be detected using the Markov Model algorithm defined in [RFC3611](#). The BurstPacketLossFrequency is

Clark

[Page 8]

Metrics Framework

9 March 2009

calculated by counting the number of burst events within the defined measurement interval. A burst that spans the boundary between two time intervals shall be counted within the later of the two intervals.

Units of Measurement: Bursts per time interval (e.g. per second, per hour, per day)

Measurement Timing: This metric can be used over a wide range of time intervals. Using time intervals of longer than one hour may prevent the detection of variations in the value of this metric due to time-of-day changes in network load. Timing intervals should not vary in duration by more than +/- 2%.

Implementation Guidelines: See [RFC3611](#).

Verification Testing: See Appendix for C code to generate test vectors.

Use and Applications: This metric is useful to detect IP network transients that affect the performance of applications such as Voice over IP or IP Video. The value of Gmin may be selected to ensure that bursts correspond to a packet loss rate that would degrade the performance of the application of interest (e.g. 16 for VoIP).

Reporting Model: This metric needs to be associated with a defined time interval, which could be defined by fixed intervals or by a sliding window.

[3.5.](#) Dependencies

[3.5.1.](#) Timing accuracy

The accuracy of the timing of a measurement may affect the accuracy of the metric. This may not materially affect a sampled value metric however would affect an interval based metric. Some metrics, for example the number of events per time interval, would be directly affected; for example a 10% variation in time interval would lead directly to a 10% variation in the measured value. Other metrics, such as the average packet loss rate during some time interval, would be affected to a lesser extent.

If it is necessary to correlate sampled values or intervals then it is essential that the accuracy of sampling time and interval start/stop times is sufficient for the application (for example +/- 2%).

[3.5.2.](#) Dependencies of metric definitions on related events or metrics

Metric definitions may explicitly or implicitly rely on factors that may not be obvious. For example, the recognition of a packet as being "lost" relies on having some method to know the packet was actually lost (e.g. RTP sequence number), and some time threshold

after which a non-received packet is declared as lost. It is

important that any such dependencies are recognized and incorporated into the metric definition.

3.5.3. Relationship between application performance and lower layer metrics

Lower layer metrics may be used to compute or infer the performance of higher layer applications, potentially using an application performance model. The accuracy of this will depend on many factors including:

- (i) The completeness of the set of metrics - i.e. are there metrics for all the input values to the application performance model?
- (ii) Correlation between input variables (being measured) and application performance
- (iii) Variability in the measured metrics and how this variability affects application performance

4. Performance Metric Development Process

4.1. New Proposals for Metrics

The following entry criteria will be considered for each proposal.

Proposals SHOULD be prepared as Internet Drafts, describing the metrics and conforming to the qualifications above as much as possible.

Proposals SHOULD be vetted by the corresponding protocol development Working Group prior to discussion by the PM Entity. This aspect of the process includes an assessment of the need for the metrics proposed and assessment of the support for their development in IETF.

Proposals SHOULD include an assessment of interaction and/or overlap with work in other Standards Development Organizations.

Proposals SHOULD specify the intended audience and users of the metrics. The development process encourages participation by members of the intended audience.

Proposals SHOULD survey the existing standards work in the area and identify additional expertise that might be consulted, or possible overlap with other standards development orgs.

Proposals SHOULD identify any security and IANA requirements. Security issues could potentially involve revealing of user identifying data or the potential misuse of active test tools. IANA considerations may involve the need for a metrics registry.

4.2. Reviewing Metrics

Each metric SHOULD be assessed according to the following list of qualifications:

- o Unambiguously defined?
- o Units of Measure Specified?
- o Measurement Interval Specified?
- o Measurement Errors Identified?
- o Repeatable?
- o Implementable?
- o Assumptions concerning underlying process?
- o Use cases?
- o Correlation with application performance/ user experience?

4.3. Proposal Approval

New work item proposals SHALL be approved using the existing IETF process.

4.4. PM Entity Interaction with other WGs

The PM Entity SHALL work in partnership with the related protocol development WG when considering an Internet Draft that specifies performance metrics for a protocol. A sufficient number of individuals with expertise must be willing to consult on the draft. If the related WG has concluded, comments on the proposal should still be sought from key RFC authors and former chairs, or from the WG mailing list if it was not closed.

Existing mailing lists SHOULD be used however a dedicated mailing list MAY be initiated if necessary to facilitate work on a draft.

In some cases, it will be appropriate to have the IETF session discussion during the related protocol WG session, to maximize visibility of the effort to that WG and expand the review.

4.5. Standards Track Performance Metrics

The PM Entity will manage the progression of PM RFCs along the Standards Track. See [[I-D.bradner-metricstest](#)]. This may include the preparation of test plans to examine different implementations of the metrics to ensure that the metric definitions are clear and unambiguous (depending on the final form of the draft above).

Clark

[Page 11]

Metrics Framework

9 March 2009

5. IANA Considerations

This document makes no request of IANA.

Note to RFC Editor: this section may be removed on publication as an RFC.

6. Security Considerations

In general, the existence of framework for performance metric development does not constitute a security issue for the Internet. Metric definitions may introduce security issues and this framework recommends that those defining metrics should identify any such risk factors.

The security considerations that apply to any active measurement of live networks are relevant here as well. See [[RFC4656](#)].

7. Acknowledgements

The authors would like to thank Al Morton, Dan Romascanu, Benoit Claise, Daryl Malas and Loki Jorgenson for their comments and contributions.

8. References

[8.1.](#) Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", [RFC 4656](#), September 2006.
- [G1000] ITU-T Recommendation G.1000. Communications Quality of Service: A framework and definitions.

[8.2.](#) Informative References

- [RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis,

"Framework for IP Performance Metrics", [RFC 2330](#),
May 1998.

[Casner] "A Fine-Grained View of High Performance Networking, NANOG
22 Conf.; <http://www.nanog.org/mtg-0105/agenda.html>", May
20-22 2001.

[I-D.bradner-metricstest]

Bradner, S. and V. Paxson, "Advancement of metrics
specifications on the IETF Standards Track",
[draft-bradner-metricstest-03](#) August 2007.

Clark

[Page 12]

Metrics Framework

9 March 2009

[I-D.ietf-ippm-framework-compagg]

Morton, A., "Framework for Metric Composition",
[draft-ietf-ippm-framework-compagg-06](#) (work in progress),
February 2008.

[I-D.ietf-ippm-spatial-composition]

Morton, A. and E. Stephan, "Spatial Composition of
Metrics", [draft-ietf-ippm-spatial-composition-06](#) (work in
progress), February 2008.

Author's Address

Alan Clark
Telchemy Incorporated
2905 Premiere Parkway, Suite 280
Duluth, Georgia 30097
USA

Phone:

Fax:

Email: alan.d.clark@telchemy.com

URI:

