Network Working Group   Scott Bradner
Internet-Draft   Harvard University
                                                     Vern Paxson
                                                            ICSI
                                                       July 2007

Advancement of metrics specifications on the IETF Standards Track

<draft-bradner-metricstest-03.txt>

1. Status of this Memo

By submitting this Internet-Draft, each author represents that any
applicable patent or other IPR claims of which he or she is aware
have been or will be disclosed, and any of which he or she becomes
aware will be disclosed, in accordance with Section 6 of 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 December 24, 2007.

2. Abstract

The Internet Standards Process [RFC2026] requires that all IETF
Standards Track specifications must have "multiple, independent, and
interoperable implementations" before they can be advanced beyond
Proposed Standard status. This document specifies the test which the
IESG will use to determine if a metrics specification document meets
these requirements. It also discusses the rationale for this
process.

3. **The Nature of the Problem**

   The Internet Standards Process [RFC2026] requires that for a IETF
   specification to advance beyond the Proposed Standard level, at least
   two genetically unrelated implementations must be shown to
   interoperate correctly with all features and options. There are two
   distinct reasons for this requirement.

   The first reason is to verify that the text of the specification is
   adequately clear and accurate. This is demonstrated by showing that
   multiple implementation efforts have used the specification to
   achieved interoperable implementations.

   The second reason is to discourage excessive options and "feature
   creep". This is accomplished by requiring interoperable
   implementation of all features, including options. If an option is
   not included in at least two different interoperable implementations,
   it is safe to assume that it has not been deemed useful and must be
   removed before the specification can advance.

   In the case of a protocol specification which specifies the "bits on
   the wire" exchanged by executing state machines, the notion of
   "interoperability" is reasonably intuitive - the implementations must
   successfully "talk to each other", exchanging "bits on the wire",
   while exercising all features and options.

   In the case of a specification for a performance metric, network
   latency for example, exactly what constitutes "interoperation" is
   less obvious. This document specifies how the IESG has decided to
   judge "metric specification interoperability" in the context of the
   IETF Standards Process.

   The aim is to ensure that the dual goals of specification clarity and
   feature evaluation have been met using an interpretation of the
   concept of metric specification interoperability that strikes a
   balance between testing complexity and practicality.

4. **On The Nature of metric specifications**

   Compared to "state machine" protocols which focus on procedural
   specifications, a metric specification describes a method of testing
   and a way to report the results of this testing. One example of such
   a metric would be a way to test and report the latency that data
   packets would incur while being sent from one network location to
   another.

   Since implementations of testing metrics are by their nature stand-
   alone and do not interact with each other, the level of the

interoperability called for in the IETF standards process cannot be
simply determined by seeing that the implementations interact
properly. Instead, we can reasonably require that different
implementations verifiably give statistically equivalent results.
Verifiable equivalence takes the place of interoperability.

5. **Discussion and Recommended Process**
In order to meet their obligations under the IETF Standards Process
the IESG must be convinced that each metric specification advanced to
Draft Standard or Internet Standard status is clearly written, that
there are the required multiple verifiably equivalent
implementations, and that all options have been implemented. There
may be multiple ways to achieve this goal; this memo documents the
way that the IESG will use to determine if the requirements have been
met.

In the context of this memo, metrics are designed to measure some
characteristic of a data network. An aim of any metric definition
should be that it should be specified in a way that can reliably
measure the specific characteristic in a repeatable way. Thus
running the test a number of times on a stable network should produce
essentially the same results. In the same way, sequentially running
different implementations of software that perform the tests
described in the metric document on a stable network, or
simultaneously on a network that may or may not be stable should
produce essentially the same results.

Following these assumptions any recommendation for the advancement of
a metric specification must be accompanied by an implementation
report, as is the case with all requests for the advancement of IETF
specifications. The implementation report must include a specific
plan to test the specific metrics in the RFC in lab or real-world
networks and reports of the tests performed with two or more
implementations of the software. The test plan should cover key
parts of the specification, speak to the accuracy required for each
aspect measured and thus define "statistically equivalent" for the
specific metrics being tested. Ideally, the test plan would co-
evolve with the development of the metric, since that's when people
have the most context in their thinking regarding the different
subtleties that can arise.

The tests MAY be sequential on the same or different hardware, with
each implementation being run after the previous one finishes, or
they MAY be run in parallel with multiple implementations each on
different hardware. It is RECOMMENDED that the tests be run not in
deterministic order, but instead by executing a randomizing algorithm
on the schedule, and interspersing tests from the several

implementations at randomly selected times until all tests of all
implementations are complete. This is a way of avoiding a biased
result in relation to the network conditions and other factors while
making the comparative tests.

All of the tests for each set should be run in the same direction
between the same two points on the same network. The tests should be
run simultaneously unless the network is stable enough to ensure that
the path the data takes through the network will not change between
tests. The tests should be run multiple times and the results
compared and the mean and variance determined. The results of the
tests using different implementations of the testing software must be
within the same variance as the results obtained from multiple
executions of the same software.

In particular, if the variance using Method A is sigma^2, then the
value yielded by Method B should fall within 2*sigma of the mean
value reported by Method A at least ### % of the time (The percentage
value will be filled in during a discussion of statistics and metrics
in the upcoming IPPM meeting). Note that if Method A and Method B
are identical, and the value they are estimating is distributed
according to a Gaussian distribution, then we would expect that
Method A's estimates would fall within 2*sigma of its mean value ###
% of the time. We allow more deviation in recognition of the many
pragmatic (and sometimes systemic) difficulties in realizing
consistent network measurement, and the fact that many quantities
related to networking have distributions quite different from
Gaussian. In addition, we explicitly recognize the IESG's
prerogative to relax the restriction of ### % within 2*sigma in light
of the particular measurement factors and difficulties, providing
these are expressed (in a communication with the relevant working
group or document author) by the IESG.

If the metric has options, all of the options must be tested in the
same way. For example, if the metric includes a list of packet sizes
that can be used, all of the listed sizes should be tested. If some
of the options are not supported or tested they must be removed from
the specification before the specification can be advanced on the
standards track.

As stated above, the measurements are made over a set of points in a
lab or real world network. The Area Director(s), in consultation
with the working group chairs (if applicable), will recommend to the
IESG their judgment as to the adequacy of the set of measurement
points in covering the range of relevant network conditions.

An implementation report is required for both the advancement from
Proposed Standard to Draft Standard and from Draft Standard to

   Internet Standard. The implementation report for advancement from
   Draft Standard to Internet Standard can be an updated version of the
   one used for the advancement from Proposed Standard to Draft
   Standard.

   The prime concern of the IESG will be that the underlying reasons for
   the verifiably equivalent implementations are met, i.e. that the text
   of the specification is clear and unambiguous, and that features of
   the specification which have not garnered support have been removed.

   The implementation report will be placed on the IETF web page along
   with the other pre-advancement implementation reports and will be
   specifically referred to in the IETF Last-Call. As with all such
   implementation reports, the determination of adequacy is made by the
   IESG upon recommendation by the Area Director(s) of the relevant IETF
   Area. This determination of adequacy can be challenged during the
   Last-Call period.

6. Security Considerations

   Good, clearly written metric specifications can be of great
   assistance in the measurement and management of the Internet and thus
   assist in the reduction of some types of security threats. Some may
   view this policy as possibly leading to a reduction in the level of
   confidence people can have in metrics specifications, but the IESG
   feels that it will adequately ensure a reasonable evaluation of the
   level of clarity and ensure that unused options can be identified and
   removed before the advancement of a specification.

7. Acknowledgements
   The basic format and some of the text for this memo came from
   RFC2438], "Advancement of MIB specifications on the IETF Standards
   Track", which provides similar guidance for the advancement of MIBs.

8. References

8.1 Normative References

   [RFC2026] "The Internet Standards Process -- Revision 3", Bradner,
   October 1996

8.2 Informative References

   [RFC2438] "Advancement of MIB specifications on the IETF Standards
   Track", O'Dell, Alvestrand, Wijnen, & Bradner, October 1998

9. Author's Addresses

    Scott Bradner
    Harvard University
    29 Oxford St.
    Cambridge MA 02138

    Email: sob@harvard.edu
    Phone: +1-617-495-3864

    Vern Paxson
    International Computer Science Institute
    1947 Center St., Suite 600,
    Berkeley, CA 94704-1198
    USA

    Email: vern@icir.org

http://www.ietf.org/ipr.

Acknowledgment

changes

02->03
   correct author list