Advanced Stream and Sampling Framework
for IP Performance Metrics (IPPM)

Abstract

   To obtain repeatable results in modern networks, test descriptions
   need an expanded stream parameter framework that also augments
   aspects specified as Type-P for test packets. This memo updates the
   IP Performance Metrics (IPPM) Framework, RFC 2330, with advanced
   considerations for measurement methodology and testing. The existing
   framework mostly assumes deterministic connectivity, and that a
   single test stream will represent the characteristics of the path
   when it is aggregated with other flows. Networks have evolved and
   test stream descriptions must evolve with them; otherwise, unexpected
   network features may dominate the measured performance. This memo
   describes new stream parameters for both network characterization and
   support of application design using IPPM metrics.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   The IETF IPPM working group first created a framework for metric
   development in [RFC2330]. This framework has stood the test of time
   and enabled development of many fundamental metrics, while only being
   updated once in a specific area [RFC5835].

   The IPPM framework [RFC2330] generally relies on several assumptions,
   one of which is not explicitly stated but assumed: lightly loaded
   paths conform to the linear "serialization delay = packet size /
   capacity" equation, and they are state-less or history-less (with
   some exceptions, e.g., firewalls are mentioned). However, this does
   not hold true for many modern network technologies, such as reactive
   paths (those with demand-driven resource allocation) and links with
   time-slotted operation. Per-flow state can be observed on test
   packet streams, and such treatment will influence network
   characterization if it is not taken into account. Flow history will
   also affect the performance of applications and be perceived by their
   users.

   Moreover, Sections 4 and 6.2 of [RFC2330] explicitly recommend
   repeatable measurement metrics and methodologies. Measurements in
   today's access networks illustrate that methodological guidelines of
   [RFC2330] must be extended to capture the reactive nature of these
   networks. There are proposed extensions to allow methodologies to
   fulfill the continuity requirement stated in Section 6.2 of
   [RFC2330], but it is impossible to guarantee they can do so.
   Practical measurements confirm that some link types exhibit distinct
   responses to repeated measurements with identical stimulus, i.e.,
   identical traffic patterns. If feasible, appropriate fine-tuning of
   measurement traffic patterns can improve measurement continuity and
   repeatability for these link types as shown in [IBD].

   This memo updates the IPPM framework [RFC2330] with advanced
   considerations for measurement methodology and testing. We note that
   the scope of IPPM work at the time of the publication of [RFC2330]
   (and during more than a decade that followed) was limited to active
   techniques or those that generate packet streams that are dedicated
   to measurement and do not monitor user traffic. This memo retains
   that same scope.

   We stress that this update of [RFC2330] does not invalidate or
   require changes to the analytic metric definitions prepared in the
   IPPM working group to date. Rather, it adds considerations for
   active measurement methodologies and expands the importance of
   existing conventions and notions in [RFC2330], such as "packets of
   Type-P".

Among the evolutionary networking changes is a phenomenon we call "reactive behavior", as defined below.

1.1. Definition: Reactive Path Behavior

Reactive path behavior will be observable by the test packet stream as a repeatable phenomenon where packet transfer performance characteristics *change* according to prior observations of the packet flow of interest (at the reactive host or link). Therefore, reactive path behavior is nominally deterministic with respect to the flow of interest. Other flows or traffic load conditions may result in additional performance-affecting reactions, but these are external to the characteristics of the flow of interest.

In practice, a sender may not have absolute control of the ingress packet stream characteristics at a reactive host or link, but this does not change the deterministic reactions present there. If we measure a path, the arrival characteristics at the reactive host/link are determined by the sending characteristics and the transfer characteristics of intervening hosts and links. Identical traffic patterns at the sending host might generate different patterns at the input of the reactive host/link due to impairments in the intermediate subpath. The reactive host/link is expected to provide a deterministic response on identical input patterns (composed of all flows, including the flow of interest).

Other than the size of the payload at the layer of interest and the header itself, packet content does not influence the measurement. Reactive behavior at the IP layer is not influenced by the TCP ports in use, for example. Therefore, the indication of reactive behavior must include the layer at which measurements are instituted.

Examples include links with Active/Inactive state detectors, and hosts or links that revise their traffic serving and forwarding rates (up or down) based on packet arrival history.

Although difficult to handle from a measurement point of view, reactive paths' entities are usually designed to improve overall network performance and user experience, for example, by making capacity available to an active user. Reactive behavior may be an artifact of solutions to allocate scarce resources according to the demands of users; thus, it is an important problem to solve for measurement and other disciplines, such as application design.

1.2. 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].

2. Scope

   The purpose of this memo is to foster repeatable measurement results
   in modern networks by highlighting the key aspects of test streams
   and packets and making them part of the IPPM framework.

   The scope is to update key sections of [RFC2330], adding
   considerations that will aid the development of new measurement
   methodologies intended for today's IP networks. Specifically, this
   memo describes useful stream parameters that complement the
   parameters discussed in Section 11.1 of [RFC2330] and the parameters
   described in Section 4.2 of [RFC3432] for periodic streams.

   The memo also provides new considerations to update the criteria for
   metrics in Section 4 of [RFC2330], the measurement methodology in
   Section 6.2 of [RFC2330], and other topics related to the quality of
   metrics and methods (see Section 4).

   Other topics in [RFC2330] that might be updated or augmented are
   deferred to future work. This includes the topics of passive and
   various forms of hybrid active/passive measurements.

3. New or Revised Stream Parameters

   There are several areas where measurement methodology definition and
   test result interpretation will benefit from an increased
   understanding of the stream characteristics and the (possibly
   unknown) network conditions that influence the measured metrics.

   1. Network treatment depends on the fullest extent on the "packet of
       Type-P" definition in [RFC2330], and has for some time.

       * State is often maintained on the per-flow basis at various
          points in the path, where "flows" are determined by IP and
          other layers. Significant treatment differences occur with
          the simplest of Type-P parameters: packet length. Use of
          multiple lengths is RECOMMENDED.

       * Payload content optimization (compression or format
          conversion) in intermediate segments breaks the convention of
          payload correspondence when correlating measurements are made
          at different points in a path.

   2. Packet history (instantaneous or recent test rate or inactivity,
       also for non-test traffic) profoundly influences measured
       performance, in addition to all the Type-P parameters described
       in [RFC2330].

   3. Access technology may change during testing. A range of transfer
       capacities and access methods may be encountered during a test
       session. When different interfaces are used, the host seeking
       access will be aware of the technology change, which
       differentiates this form of path change from other changes in
       network state. Section 14 of [RFC2330] addresses the possibility
       that a host may have more than one attachment to the network, and
       also that assessment of the measurement path (route) is valid for
       some length of time (in Sections 5 and 7 of [RFC2330]). Here, we
       combine these two considerations under the assumption that
       changes may be more frequent and possibly have greater
       consequences on performance metrics.

   4. Paths including links or nodes with time-slotted service
       opportunities represent several challenges to measurement (when
       the service time period is appreciable):

       * Random/unbiased sampling is not possible beyond one such link
          in the path.

       * The above encourages a segmented approach to end-to-end
          measurement, as described in [RFC6049] for Network
          Characterization (as defined in [RFC6703]), to understand the
          full range of delay and delay variation on the path.
          Alternatively, if application performance estimation is the
          goal (also defined in [RFC6703]), then a stream with unbiased
          or known-bias properties [RFC3432] may be sufficient.

       * Multi-modal delay variation makes central statistics
          unimportant; others must be used instead.

   Each of these topics is treated in detail below.

3.1. Test Packet Type-P

   We recommend two Type-P parameters to be added to the factors that
   have impact on path performance measurements, namely packet length
   and payload type. Carefully choosing these parameters can improve
   measurement methodologies in their continuity and repeatability when
   deployed in reactive paths.

3.1.1. Multiple Test Packet Lengths

Many instances of network characterization using IPPM metrics have
relied on a single test packet length. When testing to assess
application performance or an aggregate of traffic, benchmarking
methods have used a range of fixed lengths and frequently augmented
fixed-size tests with a mixture of sizes, or Internet Mix (IMIX) as
described in [RFC6985].

Test packet length influences delay measurements, in that the IPPM
one-way delay metric [RFC2679] includes serialization time in its
first-bit to last-bit timestamping requirements. However, different
sizes can have a larger influence on link delay and link delay
variation than serialization would explain alone. This effect can be
non-linear and change the instantaneous network performance when a
different size is used, or the performance of packets following the
size change.

Repeatability is a main measurement methodology goal as stated in
Section 6.2 of [RFC2330]. To eliminate packet length as a potential
measurement uncertainty factor, successive measurements must use
identical traffic patterns. In practice, a combination of random
payload and random start time can yield representative results as
illustrated in [IRR].

3.1.2. Test Packet Payload Content Optimization

The aim for efficient network resource use has resulted in deployment
of server-only or client-server lossless or lossy payload compression
techniques on some links or paths. These optimizers attempt to
compress high-volume traffic in order to reduce network load. Files
are analyzed by application-layer parsers, and parts (like comments)
might be dropped. Although typically acting on HTTP or JPEG files,
compression might affect measurement packets, too. In particular,
measurement packets are qualified for efficient compression when they
use standard plain-text payload. We note that use of transport-layer
encryption will counteract the deployment of network-based analysis
and may reduce the adoption of payload optimizations, however.

IPPM-conforming measurements should add packet payload content as a
Type-P parameter, which can help to improve measurement determinism.
Some packet payloads are more susceptible to compression than others,
but optimizers in the measurement path can be out ruled by using
incompressible packet payload. This payload content could be
supplied by a pseudo-random sequence generator or by using part of a
compressed file (e.g., a part of a ZIP compressed archive).

Optimization can go beyond the scope of one single data or
measurement stream. Many more client- or network-centric
optimization technologies have been proposed or standardized so far,
including Robust Header Compression (ROHC) and Voice over IP
aggregation as presented, for instance, in [EEAW]. Where
optimization is feasible and valuable, many more of these
technologies may follow. As a general observation, the more
concurrent flows an intermediate host treats and the longer the paths
shared by flows are, the higher becomes the incentive of hosts to
aggregate flows belonging to distinct sources. Measurements should
consider this potential additional source of uncertainty with respect
to repeatability. Aggregation of flows in networking devices can,
for instance, result in reciprocal timing and performance influence
of these flows, which may exceed typical reciprocal queueing effects
by orders of magnitude.

3.2. Packet History

Recent packet history and instantaneous data rate influence
measurement results for reactive links supporting on-demand capacity
allocation. Measurement uncertainty may be reduced by knowledge of
measurement packet history and total host load. Additionally, small
changes in history, e.g., because of lost packets along the path, can
be the cause of large performance variations.

For instance, delay in reactive 3G networks like High Speed Packet
Access (HSPA) depends to a large extent on the test traffic data
rate. The reactive resource allocation strategy in these networks
affects the uplink direction in particular. Small changes in data
rate can be the reason of more than a 200% increase in delay,
depending on the specific packet size. A detailed theoretical and
practical analysis of Radio Resource Control (RRC) link transitions,
which can cause such behavior in Universal Mobile Terrestrial System
(UMTS) networks, is presented, e.g., in [RRC].

3.3. Access Technology Change

[RFC2330] discussed the scenario of multi-homed hosts. If hosts
become aware of access technology changes (e.g., because of IP
address changes or lower-layer information) and make this information
available, measurement methodologies can use this information to
improve measurement representativeness and relevance.

However, today's various access network technologies can present the
same physical interface to the host. A host may or may not become
aware when its access technology changes on such an interface.
Measurements for paths that support on-demand capacity allocation
are, therefore, challenging in that it is difficult to differentiate

between access technology changes (e.g., because of mobility) and
reactive path behavior (e.g., because of data rate change).

3.4. Time-Slotted Randomness Cancellation

Time-slotted operation of path entities -- interfaces, routers, or
links -- in a network path is a particular challenge for
measurements, especially if the time-slot period is substantial. The
central observation as an extension to Poisson stream sampling in
[RFC2330] is that the first such time-slotted component cancels
unbiased measurement stream sampling. In the worst case, time-
slotted operation converts an unbiased, random measurement packet
stream into a periodic packet stream. Being heavily biased, these
packets may interact with periodic behavior of subsequent time-
slotted network entities [TSRC].

Time-slotted randomness cancellation (TSRC) sources can be found in
virtually any system, network component or path, their impact on
measurements being a matter of the order of magnitude when compared
to the metric under observation. Examples of TSRC sources include,
but are not limited to, system clock resolution, operating system
ticks, time-slotted component or network operation, etc. The amount
of measurement bias is determined by the particular measurement
stream, relative offset between allocated time slots in subsequent
path entities, delay variation in these paths, and other sources of
variation. Measurement results might change over time, depending on
how accurately the sending host, receiving host, and time-slotted
components in the measurement path are synchronized to each other and
to global time. If path segments maintain flow state, flow parameter
change or flow reallocations can cause substantial variation in
measurement results.

Practical measurements confirm that such interference limits delay
measurement variation to a subset of theoretical value range.
Measurement samples for such cases can aggregate on artificial
limits, generating multi-modal distributions as demonstrated in
[IRR]. In this context, the desirable measurement sample statistics
differentiate between multi-modal delay distributions caused by
reactive path behavior and the ones due to time-slotted interference.

Measurement methodology selection for time-slotted paths depends to a
large extent on the respective viewpoint. End-to-end metrics can
provide accurate measurement results for short-term sessions and low
likelihood of flow state modifications. Applications or services
that aim at approximating path performance for a short time interval
(in the order of minutes) and expect stable path conditions should,

therefore, prefer end-to-end metrics. Here, stable path conditions
refer to any kind of global knowledge concerning measurement path
flow state and flow parameters.

However, if long-term forecast of time-slotted path performance is
the main measurement goal, a segmented approach relying on
measurement of subpath metrics is preferred. Regenerating unbiased
measurement traffic at any hop can help to reveal the true range of
path performance for all path segments.

4. Quality of Metrics and Methodologies

[RFC6808] proposes repeatability and continuity as one of the metric
and methodology properties to infer on measurement quality.
Depending mainly on the set of controlled measurement parameters,
measurements repeated for a specific network path using a specific
methodology may or may not yield repeatable results. Challenging
measurement scenarios for adequate parameter control include
wireless, reactive, or time-slotted networks as discussed earlier in
this document. This section presents an expanded definition of
"repeatability" beyond the definition in [RFC2330] and an expanded
examination of the concept of "continuity" in [RFC2330] and its
limited applicability.

4.1. Revised Definition of Repeatability

[RFC2330] defines repeatability in a general way:

"A methodology for a metric should have the property that it is
repeatable: if the methodology is used multiple times under identical
conditions, the same measurements should result in the same
measurements."

The challenge is to develop this definition further, such that it
becomes an objective measurable criterion (and does not depend on the
concept of continuity discussed below). Fortunately, this topic has
been treated in other IPPM work. In BCP 176 [RFC6576], the criteria
of equivalent results was agreed as the surrogate for
interoperability when assessing metric RFCs for Standards Track
advancement. The criteria of equivalence were expressed as objective
statistical requirements for comparison across the same
implementations and independent implementations in the test plans
specific to each RFC evaluated ([RFC2679] in the test plan of
[RFC6808]).

The tests of [RFC6808] rely on nearly identical conditions to be
present for analysis and accept that these conditions cannot be
exactly identical in the production network paths used. The test

plans allow some correction factors to be applied (some statistical
tests are hyper-sensitive to differences in the mean of
distributions) and recognize the original findings of [RFC2330]
regarding excess sample sizes.

One way to view the reliance on identical conditions is to view it as
a challenge: How few parameters and path conditions need to be
controlled and still produce repeatable methods/measurements?

Although the test plan in [RFC6808] documented numerical criteria for
equivalence, we cannot specify the exact numerical criteria for
repeatability *in general*. The process in the BCP [RFC6576] and
statistics in [RFC6808] have been used successfully, and the
numerical criteria to declare a metric repeatable should be agreed by
all interested parties prior to measurement.

We revise the definition slightly, as follows:

    A methodology for a metric should have the property that it is
    repeatable: if the methodology is used multiple times under
    identical conditions, the methods should produce equivalent
    measurement results.

4.2. Continuity No Longer an Alternative Repeatability Criterion

In the original framework [RFC2330], the concept of continuity was
introduced to provide a relaxed criteria for judging repeatability
and was described in Section 6.2 of [RFC2330] as follows:

"...a methodology for a given metric exhibits continuity if, for
small variations in conditions, it results in small variations in the
resulting measurements."

Although there are conditions where metrics may exhibit continuity,
there are others where this criteria would fail for both user traffic
and active measurement traffic. Consider link fragmentation and the
non-linear increase in delay when we increase packet size just beyond
the limit of a single fragment. An active measurement packet would
see the same delay increase when exceeding the fragment size.

The Bulk Transfer Capacity (BTC) [RFC3148] gives another example in
Section 1, bottom of page 2:

    There is also evidence that most TCP implementations exhibit non-
    linear performance over some portion of their operating region.
    It is possible to construct simple simulation examples where
    incremental improvements to a path (such as raising the link data
    rate) results in lower overall TCP throughput (or BTC) [Mat98].

Clearly, the time-slotted network elements described in Section 3.4
of this document also qualify as a new exception to the ideal of
continuity.

   Therefore, we deprecate continuity as an alternate criterion on
   metrics and prefer the more exact evaluation of repeatability
   instead.

4.3. Metrics Should Be Actionable

   The IP Performance Metrics Framework [RFC2330] includes usefulness as
   a metric criterion:

   "...The metrics must be useful to users and providers in
   understanding the performance they experience or provide...".

   When considering measurements as part of a maintenance process,
   evaluation of measurement results for a path under observation can
   draw attention to potential performance problems "somewhere" on the
   path. Anomaly detection is, therefore, an important phase and first
   step that already satisfies the usefulness criterion for many
   metrics.

   This concept of usefulness can be extended, becoming a subset of what
   we refer to as "actionable" criterion in the following. We note that
   this is not the term from law.

   Central to maintenance is the isolation of the root cause of reported
   anomalies down to a specific subpath, link or host, and metrics
   should support this second step as well. While detection of path
   anomaly may be the result of an on-going monitoring process, the
   second step of cause isolation consists of specific, directed on-
   demand measurements on components and subpaths. Metrics must support
   users in this directed search, becoming actionable:

      Metrics must enable users and operators to understand path
      performance and SHOULD help to direct corrective actions when
      warranted, based on the measurement results.

   Besides characterizing metrics, usefulness and actionable properties
   are also applicable to methodologies and measurements.

4.4. It May Not Be Possible To Be Conservative

   [RFC2330] adopts the term "conservative" for measurement
   methodologies for which:

   "... the act of measurement does not modify, or only slightly
   modifies, the value of the performance metric the methodology
   attempts to measure."

   It should be noted that this definition of "conservative" in the
   sense of [RFC2330] depends to a large extent on the measurement
   path's technology and characteristics. In particular, when deployed
   on reactive paths, subpaths, links or hosts conforming to the
   definition in Section 1.1 of this document, measurement packets can
   originate capacity (re)allocations. In addition, small measurement
   flow variations can result in other users on the same path perceiving
   significant variations in measurement results. Therefore:

      It is not always possible for the method to be conservative.

4.5. Spatial and Temporal Composition Support Unbiased Sampling

   Concepts related to temporal and spatial composition of metrics in
   Section 9 of [RFC2330] have been extended in [RFC5835]. [RFC5835]
   defines multiple new types of metrics, including Spatial Composition,
   Temporal Aggregation, and Spatial Aggregation. So far, only the
   metrics for Spatial Composition have been standardized [RFC6049],
   providing the ability to estimate the performance of a complete path
   from subpath metrics. Spatial Composition aligns with the finding of
   [TSRC] that unbiased sampling is not possible beyond the first time-
   slotted link within a measurement path.

      In cases where unbiased measurement for all segments of a path is
      not feasible due to the presence of a time-slotted link, restoring
      randomness of measurement samples when necessary is recommended as
      presented in [TSRC], in combination with Spatial Composition
      [RFC6049].

4.6. When to Truncate the Poisson Sampling Distribution

   Section 11.1.1 of [RFC2330] describes Poisson sampling, where the
   inter-packet send times have a Poisson distribution. A path element
   with reactive behavior sensitive to flow inactivity could change
   state if the random inter-packet time is too long.

      It is recommended to truncate the tail of Poisson distribution
      when needed to avoid reactive element state changes.

Tail truncation has been used without issue to ensure that minimum
sample sizes can be attained in a fixed-test interval.

5. Conclusions

Safeguarding repeatability as a key property of measurement
methodologies is highly challenging and sometimes impossible in
reactive paths. Measurements in paths with demand-driven allocation
strategies must use a prototypical application packet stream to infer
a specific application's performance. Measurement repetition with
unbiased network and flow states (e.g., by rebooting measurement
hosts) can help to avoid interference with periodic network behavior,
with randomness being a mandatory feature for avoiding correlation
with network timing.

Inferring the path performance between one measurement session or
packet stream and other sessions/streams with alternate
characteristics is generally discouraged with reactive paths because
of the huge set of global parameters that have influence on
instantaneous path performance.

6. Security Considerations

The security considerations that apply to any active measurement of
live paths are relevant here as well. See [RFC4656] and [RFC5357].

When considering privacy of those involved in measurement or those
whose traffic is measured, the sensitive information available to
potential observers is greatly reduced when using active techniques
that are within this scope of work. Passive observations of user
traffic for measurement purposes raise many privacy issues. We refer
the reader to the privacy considerations described in the Large Scale
Measurement of Broadband Performance (LMAP) Framework [LMAP], which
covers active and passive techniques.

7. Acknowledgements

The authors thank Rudiger Geib, Matt Mathis, Konstantinos
Pentikousis, and Robert Sparks for their helpful comments on this
memo, Alissa Cooper and Kathleen Moriarty for suggesting ways to
"update the update" for heightened privacy awareness and its
consequences, and Ann Cerveny for her editorial review and comments
that helped to improve readability overall.

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.

[RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis,
           "Framework for IP Performance Metrics", RFC 2330, May
           1998.

[RFC2679] Almes, G., Kalidindi, S., and M. Zekauskas, "A One-way
           Delay Metric for IPPM", RFC 2679, September 1999.

[RFC3432] Raisanen, V., Grotefeld, G., and A. Morton, "Network
           performance measurement with periodic streams", RFC 3432,
           November 2002.

[RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M.
           Zekauskas, "A One-way Active Measurement Protocol
           (OWAMP)", RFC 4656, September 2006.

[RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J.
           Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)",
           RFC 5357, October 2008.

[RFC5835] Morton, A. and S. Van den Berghe, "Framework for Metric
           Composition", RFC 5835, April 2010.

[RFC6049] Morton, A. and E. Stephan, "Spatial Composition of
           Metrics", RFC 6049, January 2011.

[RFC6576] Geib, R., Morton, A., Fardid, R., and A. Steinmitz, "IP
           Performance Metrics (IPPM) Standard Advancement Testing",
           BCP 176, RFC 6576, March 2012.

[RFC6703] Morton, A., Ramachandran, G., and G. Maguluri, "Reporting
           IP Network Performance Metrics: Different Points of View",
           RFC 6703, August 2012.

8.2. Informative References

   [EEAW]   Pentikousis, K., Piri, E., Pinola, J., Fitzek, F.,
               Nissilae, T., and I. Harjula, "Empirical Evaluation of
               VoIP Aggregation over a Fixed WiMAX Testbed", Proceedings
               of the 4th International Conference on Testbeds and
               research infrastructures for the development of networks
               and communities (TridentCom '08), Article No. 19, March
               2008, <http://dl.acm.org/citation.cfm?id=139059>.

   [IBD]   Fabini, J., Karner, W., Wallentin, L., and T. Baumgartner,
               "The Illusion of Being Deterministic - Application-Level
               Considerations on Delay in 3G HSPA Networks", Lecture
               Notes in Computer Science, Volume 5550, pp. 301-312 , May
               2009.

   [IRR]   Fabini, J., Wallentin, L., and P. Reichl, "The Importance
               of Being Really Random: Methodological Aspects of IP-Layer
               2G and 3G Network Delay Assessment", ICC'09 Proceedings of
               the 2009 IEEE International Conference on Communications,
               doi: 10.1109/ICC.2009.5199514, June 2009.

   [LMAP]   Eardley, P., Morton, A., Bagnulo, M., Burbridge, T.,
               Aitken, P., and A. Akhter, "A framework for large-scale
               measurement platforms (LMAP)", Work in Progress, June
               2014.

   [Mat98]   Mathis, M., "Empirical Bulk Transfer Capacity", IP
               Performance Metrics Working Group report in Proceedings of
               the Forty-Third Internet Engineering Task Force, Orlando,
               FL, December 1998,
               <http://www.ietf.org/proceedings/43/slides/
               ippm-mathis-98dec.pdf>.

   [RFC3148]   Mathis, M. and M. Allman, "A Framework for Defining
               Empirical Bulk Transfer Capacity Metrics", RFC 3148, July
               2001.

   [RFC6808]   Ciavattone, L., Geib, R., Morton, A., and M. Wieser, "Test
               Plan and Results Supporting Advancement of RFC 2679 on the
               Standards Track", RFC 6808, December 2012.

   [RFC6985]   Morton, A., "IMIX Genome: Specification of Variable Packet
               Sizes for Additional Testing", RFC 6985, July 2013.

   [RRC]   Peraelae, P., Barbuzzi, A., Boggia, G., and K.
               Pentikousis, "Theory and Practice of RRC State Transitions
               in UMTS Networks", IEEE Globecom 2009 Workshops, doi:
               10.1109/GLOCOMW.2009.5360763, November 2009.

   [TSRC]   Fabini, J. and M. Abmayer, "Delay Measurement Methodology
               Revisited: Time-slotted Randomness Cancellation", IEEE
               Transactions on Instrumentation and Measurement, Volume
               62, Issue 10, doi:10.1109/TIM.2013.2263914, October 2013.

Authors' Addresses

   Joachim Fabini
   Vienna University of Technology
   Gusshausstrasse 25/E389
   Vienna 1040
   Austria

   Phone: +43 1 58801 38813
   Fax: +43 1 58801 38898
   EMail: Joachim.Fabini@tuwien.ac.at
   URI: http://www.tc.tuwien.ac.at/about-us/staff/joachim-fabini/

   Al Morton
   AT&T Labs
   200 Laurel Avenue South
   Middletown, NJ 07748
   USA

   Phone: +1 732 420 1571
   Fax: +1 732 368 1192
   EMail: acmorton@att.com
   URI: http://home.comcast.net/~acmacm/