Network Working Group	A. Morton, Ed.
Internet-Draft	AT&T Labs
Intended status: Informational	S. Van den Berghe, Ed.
Expires: May 8, 2008	Ghent University - IBBT
	November 05, 2007

Framework for Metric Composition
draft-ietf-ippm-framework-compagg-05

Status of this Memo

Abstract

This memo describes a framework for composing and aggregating metrics (both in time and in space) defined by RFC 2330 and developed by the IPPM working group. The framework describes the generic composition and aggregation mechanisms. It provides a basis for additional documents that implement this framework for detailed, and practically useful, compositions and aggregations of metrics.

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 (Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” March 1997.) [RFC2119].

Table of Contents

---

1. Introduction

TOC

The IPPM framework [RFC2330] (Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, “Framework for IP Performance Metrics,” May 1998.) describes two forms of metric composition, spatial and temporal. Also, the text suggests that the concepts of the analytical framework (or A-frame) would help to develop useful relationships to derive the composed metrics from real metrics. The effectiveness of composed metrics is dependent on their usefulness in analysis and applicability to practical measurement circumstances.

This memo expands on the notion of composition, and provides a detailed framework for several classes of metrics that were mentioned in the original IPPM framework. The classes include temporal aggregation, spatial aggregation, and spatial composition.

1.1. Motivation

TOC

Network operators have deployed measurement systems to serve many purposes, including performance monitoring, maintenance support, network engineering, and customer reporting. The collection of elementary measurements alone is not enough to understand a network's behaviour. In general, measurements need to be post-processed to present the most relevant information for each purpose. The first step is often a process of "composition" of single measurements or measurement sets into other forms. Composition and aggregation present several more post-processing opportunities to the network operator, and we describe the key motivations below.

1.1.1. Reducing Measurement Overhead

TOC

A network's measurement possibilities scale upward with the square of the number of nodes. But each measurement implies overhead, in terms of the storage for the results, the traffic on the network (assuming active methods), and the OA&M for the measurement system itself. In a large network, it is impossible to perform measurements from each node to all others.

An individual network operator should be able to organize their measurement paths along the lines of physical topology, or routing areas/Autonomous Systems, and thus minimize dependencies and overlap between different measurement paths. This way, the sheer number of measurements can be reduced, as long as the operator has a set of methods to estimate performance between any particular nodes when needed.

Composition and aggregation play a key role when the path of interest spans multiple networks, and where each operator conducts their own measurements. Here, the complete path performance may be estimated from measurements on the component parts.
Operators that take advantage of the composition and aggregation methods recognize that the estimates may exhibit some additional error beyond that inherent in the measurements themselves, and so they are making a trade-off to achieve reasonable measurement system overhead.

1.1.2. Measurement Re-use

TOC

There are many different measurement users, each bringing specific requirements for the reporting timescale. Network managers and maintenance forces prefer to see results presented very rapidly, to detect problems quickly or see if their action has corrected a problem. On the other hand, network capacity planners and even network users sometimes prefer a long-term view of performance, for example to check trends. How can one set of measurements serve both needs?
The answer lies in temporal aggregation, where the short-term measurements needed by the operations community are combined to estimate a longer-term result for others. Also, problems with the measurement system itself may be isolated to one or more of the short-term measurements, rather than possibly invalidating an entire long-term measurement if the problem was undetected.

1.1.3. Data Reduction and Consolidation

TOC

Another motivation is data reduction. Assume there is a network domain in which delay measurements are performed among a subset of its nodes. A network manager might ask whether there is a problem with the network delay in general. It would be desirable to obtain a single value that gives an indication of the overall network delay. Spatial aggregation methods would address this need, and can produce the desired "single figure of merit" asked for, one that may also be useful in trend analysis.
The overall value would be calculated from the elementary delay measurements, but it not obvious how: for example, it may not to be reasonable to average all delay measurements, as some paths (e.g. having a higher bandwidth or more important customers) might be considered more critical than others.
Metric composition can help to provide, from raw measurement data, some tangible, well-understood and agreed upon information about the service guarantees provided by a network. Such information can be used in the

Service Level Agreement/Service Level Specification (SLA/SLS) contracts between a service provider and its customers.

1.1.4. Implications on Measurement Design and Reporting

TOC

If a network measurement system operator anticipates needing to produce overall metrics by composition, then it is prudent to keep that requirement in mind when considering the measurement design and sampling plan. Also, certain summary statistics are more conducive to composition than others, and this figures prominently in the design of measurements and when reporting the results.

---

2. Purpose and Scope

TOC

The purpose of this memo is provide a common framework for the various classes of metrics based on composition of primary metrics. The scope is limited to the definitions of metrics that are composed from primary metrics using a deterministic function. Key information about each metric, such as the assumptions under which the relationship holds and possible sources of error/circumstances where the composition may fail, are included.

At this time, the scope of effort is limited to the metrics for packet loss, delay, and delay variation. Composition of packet reordering metrics is considered a research topic at the time this memo was prepared, and beyond its scope.

This memo will retain the terminology of the IPPM Framework [RFC2330] (Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, “Framework for IP Performance Metrics,” May 1998.)as much as possible, but will extend the terminology when necessary. It is assumed that the reader is familiar with the concepts introduced in [RFC2330] (Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, “Framework for IP Performance Metrics,” May 1998.), as they will not be repeated here.

---

3. Terminology

TOC

This section defines the terminology applicable to the processes of Metric Composition and Aggregation.

TOC

3.1. Measurement Point

The logical or physical location where packet observations are made. The term Measurement Point is synonymous with the term "observation position" used in [RFC2330] (Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, “Framework for IP Performance Metrics,” May 1998.) when describing the notion of wire time. A measurement point may be at the boundary between a host and an adjacent link (physical), or it may be within a host (logical) that performs measurements where the difference between host time and wire time is understood.

3.2. Complete path

TOC

The complete path is the true path that a packet would follow as it traverses from the packet’s Source to its Destination.

3.3. Complete path metric

TOC

The complete path metric is the Source to Destination metric that a composed metric is estimating. A complete path metric represents the ground-truth for a composed metric.

3.4. Composed Metric

TOC

A composed metric is an estimate of an actual metric describing the performance of a path over some time interval. A composed metric is derived from other metrics by applying a deterministic process or function (e.g., a composition function). The process may use metrics that are identical to the metric being composed, or metrics that are dissimilar, or some combination of both types.

3.5. Composition Function

TOC

A composition function is a deterministic process applied to individual metrics to derive another metric (such as a Composed metric).

TOC

3.6. Index

An Index is a composed 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. The composition function for an index is often developed after the index range and index behavior have been determined. Examples include the R factor, as described in [G.107] (ITU-T Recommendation G.107, “"The E-model, a computational model for use in transmission planning",” March 2005.).

TOC

3.7. Ground Truth

As applied here, the notion of ground truth is defined as the actual performance of a network path over some time interval. The ground truth is metric based on the (unavailable) measurement that a composed metric seeks to estimate.

TOC

3.8. Sub-interval

A Sub-interval is a time interval that is included in another interval.

TOC

3.9. Sub-path

A Sub-path is a portion of the complete path where at least the Sub-path Source and Destination hosts are constituents of the complete path. We say that this sub-path is “involved” in the complete path.

TOC

3.10. Sub-path metrics

A sub-path path metric is an element of the process to derive a Composite metric, quantifying some aspect of the performance a particular sub-path from its Source to Destination.

TOC

4. Description of Metric Types

This section defines the various classes of Composition. There are two classes more accurately described as aggregation over time and space, and the third involves concatenation in space.

4.1. Temporal Aggregation Description

TOC

Aggregation in time is defined as the composition of metrics with the same type and scope obtained in different time instants or time windows. For example, starting from a time series of the measurements of maximum and minimum One-Way Delay on a certain network path obtained over 5-minute intervals, we obtain a time series measurement with a coarser resolution (60 minutes) by taking the max of 12 consecutive 5-minute maxima and the min of 12 consecutive 5-minute minima.
The main reason for doing time aggregation is to reduce the amount of data that has to be stored, and make the visualization/spotting of regular cycles and/or growing or decreasing trends easier. Another useful application is to detect anomalies or abnormal changes in the network characteristics.
In RFC 2330, the term "temporal composition" is introduced and differs from temporal aggregation in that it refers to methodologies to predict future metrics on the basis of past observations, exploiting the time correlation that certain metrics can exhibit. We do not consider this type of composition here.
>>>>>>>>Comment: Why no forecasting? This was apparently a limit on the Geant2 project, but may not apply here.

4.2. Spatial Aggregation Description

TOC

Aggregation in space is defined as the combination of metrics of the same type and different scope, in order to estimate the overall performance of a larger domain. This combination may involve weighing the contributions of the input metrics.
Suppose we want to compose the average One-Way-Delay (OWD) experienced by flows traversing all the Origin-Destination (OD) pairs of a network domain (where the inputs are already metric "statistics"). Since we wish to include the effect of the traffic matrix on the result, it makes sense to weight each metric according to the traffic carried on the corresponding OD pair:
OWD_sum=f1*OWD_1+f2*OWD_2+...+fn*OWD_n
where fi=load_OD_i/total_load.
A simple average OWD across all network OD pairs would not use the traffic weighting.

Another example metric that is "aggregated in space", is the maximum edge-to-edge delay across a single domain. Assume that a Service Provider wants to advertise the maximum delay that transit traffic will experience while passing through his/her domain. There can be multiple edge-to-edge paths across a domain, and the Service Provider chooses either to publish a list of delays (each corresponding to a specific edge-to-edge path), or publish a single maximum value. The latter approach simplifies the publication of measurement information, and may be sufficient for some purposes. Similar operations can be provided to other metrics, e.g. "maximum edge-to-edge packet loss", etc.
We suggest that space aggregation is generally useful to obtain a summary view of the behaviour of large network portions, or in general of coarser aggregates. The metric collection time instant, i.e. the metric collection time window of measured metrics is not considered in space aggregation. We assume that either it is consistent for all the composed metrics, e.g. compose a set of average delays all referred to the same time window, or the time window of each composed metric does not affect aggregated metric.

4.3. Spatial Composition Description

TOC

Concatenation in space is defined as the composition of metrics of same type and (ideally) different spatial scope, so that the resulting metric is representative of what the metric would be if obtained with a direct measurement over the sequence of the several spatial scopes. An example is the sum of OWDs of different edge-to-edge domain's delays, where the intermediate edge points are close to each other or happen to be the same. In this way, we can for example estimate OWD_AC starting from the knowledge of OWD_AB and OWD_BC. Note that there may be small gaps in measurement coverage, likewise there may be small overlaps (e.g., the link where test equipment connects to the network).
One key difference from examples of aggregation in space is that all sub-paths contribute equally to the composed metric, independent of the traffic load present.

4.4. Help Metrics

TOC

Finally, note that in practice there is often the need of extracting a new metric making some computation over one or more metrics with the same spatial and time scope. For example, the composed metric rtt_sample_variance may be composed from two different metrics: the help metric rtt_square_sum and the statistical metric rtt_sum. This operation is however more a simple calculation and not an aggregation or a concatenation, and we'll not investigate it further in this memo.

4.5. Higher Order Composition

TOC

Composed metrics might themselves be subject to further steps of composition or aggregation. An example would be a the delay of a maximal domain obtained through the spatial composition of several composed end-to-end delays (obtained through spatial composition). All requirements for first order composition metrics apply to higher order composition.
>>>>> Comment Response: are more examples needed here?

5. Requirements for Composed Metrics

TOC

The definitions for all composed metrics MUST include sections to treat the following topics.
The description of each metric will clearly state:

1. the definition (and statistic, where appropriate);
2. the composition or aggregation relationship;
3. the specific conjecture on which the relationship is based and assumptions of the statistical model of the process being measured, if any (see [RFC2330] (Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, "Framework for IP Performance Metrics," May 1998.) section 12);
4. a justification of practical utility or usefulness for analysis using the A-frame concepts;
5. one or more examples of how the conjecture could be incorrect and lead to inaccuracy;
6. the information to be reported.

For each metric, the applicable circumstances will be defined, in terms of whether the composition or aggregation:

- Requires homogeneity of measurement methodologies, or can allow a degree of flexibility (e.g., active or passive methods produce the "same" metric). Also, the applicable sending streams will be specified, such as Poisson, Periodic, or both.
- Needs information or access that will only be available within an operator's domain, or is applicable to Inter-domain composition.

*Requires precisely synchronized measurement time intervals in all component metrics, or loosely synchronized, or no timing requirements.

*Requires assumption of component metric independence w.r.t. the metric being defined/composed, or other assumptions.

*Has known sources of inaccuracy/error, and identifies the sources.

---

6. Guidelines for Defining Composed Metrics

TOC

6.1. Ground Truth: Comparison with other IPPM Metrics

TOC

Figure 1 illustrates the process to derive a metric using spatial composition, and compares the composed metric to other IPPM metrics. Metrics <M1, M2, M3> describe the performance of sub-paths between the Source and Destination of interest during time interval <T, Tf>. These metrics are the inputs for a Composition Function that produces a Composed Metric.

```
                  Sub-Path Metrics
         ++  M1   ++ ++  M2   ++ ++  M3   ++
     Src ||.......|| ||.......|| ||.......|| Dst
         ++    `.  ++ ++   |   ++ ++  .'   ++
                 `.        |        .-'
                   `-.     |      .'
                      `._..|.._.'
                    ,-'         `-.
                  ,'               `.
                  |   Composition   |
                  \     Function    '
                   `._            _,'
                      `--.......--'
                           |
         ++                |              ++
     Src ||...............................|| Dst
         ++         Composed Metric       ++

         ++       Complete Path Metric    ++
     Src ||...............................|| Dst
         ++                               ++
                    Spatial Metric
         ++   S1    ++    S2    ++    S3    ++
     Src ||.........||.........||..........|| Dst
         ++         ++          ++          ++
```

Figure 1: Comparison with other IPPM metrics

The Composed Metric is an estimate of an actual metric collected over the complete Source to Destination path. We say that the Complete Path Metric represents the "Ground Truth" for the Composed Metric. In other words, Composed Metrics seek to minimize error w.r.t. the Complete Path Metric.

Further, we observe that a Spatial Metric I-D.ietf-ippm-multimetrics (Stephan, E., Liang, L., and A. Morton, "IP Performance Metrics (IPPM) for spatial and multicast," September 2009.) [I-D.ietf-ippm-multimetrics]collected for packets traveling over the same set of sub-paths provide a basis for the Ground Truth of the individual Sub-Path metrics. We note that mathematical operations may be necessary to isolate the performance of each sub-path.

Next, we consider multiparty metrics as defined in [I-D.ietf-ippm-multimetrics], and their spatial composition. Measurements to each of the Receivers produce an element of the one-to-group metric. These elements can be composed from sub-path metrics and the composed metrics can be combined to create a composed one-to-group metric. Figure 2 illustrates this process.

```
             Sub-Path Metrics
    ++  M1   ++ ++  M2   ++ ++  M3   ++
Src ||.......|| ||.......|| ||.......||Rcvr1
    ++       ++ ++`.     ++ ++       ++
                    `-.
                     M4`.++ ++  M5   ++
                         || ||.......||Rcvr2
                         ++ ++`.     ++
                                `-.
                                 M6`.++
                                     ||Rcvr3
                                     ++

            One-to-Group Metric
    ++        ++         ++          ++
Src ||........||.........||..........||Rcvr1
    ++        ++.        ++          ++
                 `-.
                    `-.  ++          ++
                       `-||..........||Rcvr2
                         ++.         ++
                            `-.
                               `-.   ++
                                  `-.||Rcvr3
                                     ++
```

Figure 2: Composition of One-to-Group Metrics

Here, Sub-path Metrics M1, M2, and M3 are combined using a relationship to compose the metric applicable to the Src-Rcvr1 path. Similarly, M1, M4, and M5 are used to compose the Src-Rcvr2 metric and M1, M4, and M6 compose the Src-Rcvr3 metric.

The Composed One-to-Group Metric would list the Src-Rcvr metrics for each Receiver in the Group:

(Composed-Rcvr1, Composed-Rcvr2, Composed-Rcvr3)

The "Ground Truth" for this composed metric is of course an actual One-to-Group metric, where a single source packet has been measured after traversing the Complete Paths to the various receivers.

TOC

6.1.1. Ground Truth for Temporal Aggregation

Temporal Aggregation involves measurements made over sub-intervals of the desired test interval between the same Source and Destination. Therefore, the "Ground Truth" is the metric measured over the desired interval.

6.1.2. Ground Truth for Spatial Aggregation

TOC

Spatial Aggregation combines many measurements using a weighting function to provide the same emphasis as though the measurements were based on actual traffic, with inherent weights. Therefore, the "Ground Truth" is the metric measured on the actual traffic instead of the active streams that sample the performance.

6.2. Deviation from the Ground Truth

TOC

A metric composition can deviate from the ground truth for several reasons. Two main aspects are:

* The propagation of the inaccuracies of the underlying measurements when composing the metric. As part of the composition function, errors of measurements might propagate. Where possible, this analysis should be made and included with the description of each metric.

* A difference in scope. When concatenating hop-by-hop active measurement results to obtain the end-to-end metric, the actual measured path will not be identical to the end-to-end path. It is in general difficult to quantify this deviation, but a metric definition might identify guidelines for keeping the deviation as small as possible.

The description of the metric composition MUST include an section identifying the deviation from the ground truth.

6.3. Incomplete Information

TOC

In practice, when measurements cannot be initiated on a sub-path or during a particular measurement interval (and perhaps the measurement system gives up during the test interval), then there will not be a value for the subpath reported, and the result SHOULD be recorded as "undefined".

---

7. IANA Considerations

TOC

This document makes no request of IANA.
Note to RFC Editor: this section may be removed on publication as an RFC.

---

8. Security Considerations

TOC

The security considerations that apply to any active measurement of live networks are relevant here as well. See [RFC4656] (Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, “A One-way Active Measurement Protocol (OWAMP),” September 2006.).

---

9. Acknowledgements

TOC

The authors would like to thank Maurizio Molina, Andy Van Maele, Andreas Haneman, Igor Velimirovic, Andreas Solberg, Athanassios Liakopulos, David Schitz, Nicolas Simar and the Geant2 Project. We also acknowledge comments and suggestions from Phil Chimento, Emile Stephan, Lei Liang, Stephen Wolff, and Alan Clark.

---

10. References

TOC

10.1. Normative References

TOC

[I-D.ietf-ippm-multimetrics]	Stephan, E., Liang, L., and A. Morton, “IP Performance Metrics (IPPM) for spatial and multicast,” draft-ietf-ippm-multimetrics-12 (work in progress), September 2009 (TXT).
[RFC2119]	Bradner, S., “Key words for use in RFCs to Indicate Requirement Levels,” BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).
[RFC2330]	

	Paxson, V., Almes, G., Mahdavi, J., and M. Mathis, “Framework for IP Performance Metrics,” RFC 2330, May 1998 (TXT, HTML, XML).
[RFC4656]	Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, “A One-way Active Measurement Protocol (OWAMP),” RFC 4656, September 2006 (TXT).

10.2. Informative References

TOC

[G. 107]	ITU-T Recommendation G.107, “"The E-model, a computational model for use in transmission planning",” March 2005.

Authors' Addresses

TOC

	Al Morton (editor)
	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/
	Steven Van den Berghe (editor)
	Ghent University - IBBT
	G. Crommenlaan 8 bus 201
	Gent 9050
	Belgium
Phone:	+32 9 331 49 73
Email:	steven.vandenberghe@intec.ugent.be
URI:	http://www.ibcn.intec.ugent.be

Full Copyright Statement

TOC