International Telecommunication Union

ITU-T

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

Y.1543

(11/2007)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

Internet protocol aspects – Quality of service and network performance

---

Measurements in IP networks for inter-domain performance assessment

Recommendation ITU-T Y.1543

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	Y.1200–Y.1299
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	**Y.1500–Y.1599**
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Numbering, naming and addressing	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.1543

Measurements in IP networks for inter-domain performance assessment

Summary

Recommendation ITU-T Y.1543 specifies a set of IP performance parameters and methods of measurement applicable when assessing the quality of packet transfer on inter-domain paths. The methods anticipate that there will be multiple measurement systems, each conducting measurements of a segment of the customer-to-customer path, and recommend configurations that should produce useful results in this cooperative scenario. The methods rely on existing parameter definitions and encompass both active and passive measurement techniques.

Source

Recommendation ITU-T Y.1543 was approved on 13 November 2007 by ITU-T Study Group 12 (2005-2008) under Recommendation ITU-T A.8 procedure.

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g. interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had not received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Introduction

Network performance expectations must be set and monitored among users and service providers to raise confidence in network delivery. Users typically only see the end-to-end performance, i.e., the concatenation of performance over multiple network segments and/or across multiple heterogeneous service providers. Thus, meaningful discussions of network quality of service (QoS) between users and service providers are most relevant on an end-to-end basis.

Existing standards specify several metrics and measurement methods for point-to-point performance. Notable are Recommendations ITU-T Y.1540 and Y.1541 and the IETF IP Performance Metrics (IPPM) Working Group standards. However, many options and parameters are left unspecified, as are mapping between IP and non-IP metrics, accuracy and data handling. Each of these topics must be specified in order to support QoS across multiple heterogeneous service providers. Therefore, this Recommendation specifies the essential measurement options, so that performance measurements conducted by operators in their administrative domains can be easily combined to estimate the end-to-end network performance or the inter-domain QoS.

Recommendation ITU-T Y.1543

Measurements in IP networks for inter-domain performance assessment

1 Scope

This Recommendation describes measurements which are applicable for:

1) Providers' delivery assurance of customers' network performance.
2) Providers to supply performance information for prospective customers.
3) Providers' troubleshooting among their networks along defined paths.
4) Providers' internal indication of performance impact of changes within their networks.
5) Providers' monitoring of each others network performance.
6) Providing information to other NGN components, e.g., RACF, bandwidth broker, OSS/BSS, etc.

The scope of this Recommendation covers active and passive measurement and combinations of these two techniques. Active measurement employs packets dedicated to the measurement function inserted at one measurement point and collected at a remote measurement point. Passive measurement usually involves observations of user packet traffic at one or more measurement points. Spatial measurement is a special category of active measurement that employs both active and passive techniques. It utilizes observations of measurement-dedicated packets at three or more measurement points, where one or more point(s) simply monitor(s) (and do(es) not terminate) the test packets.

This Recommendation presents requirements for performance measurements including performance attributes and time-escales. Building upon existing standards, it recommends best practice in these areas based on [ITU-T Y.1540]. Comparisons to other standards (IETF RFCs) are included. Defining the probe packet format is beyond the scope of this Recommendation.

This Recommendation describes a network model which locates key points of demarcation and measurement. It categorizes various measurements and shows how they may be applied to the network model. It reviews time synchronization and sets targets for equipment which is located at various points in the network model.

Security requirements of the measurement traffic are analysed, approaches are considered, and then a set of approaches are picked. Security of BGP, synchronization systems and customer equipment are beyond the scope of this Recommendation.

Customer interactions with their service provider are discussed at a high level. Details of transferring results to customers are beyond the scope of this Recommendation.

The target networks of this Recommendation are IP networks and MPLS-enabled IP networks, pure L2 and other non-native IP networks are out of its scope.

This Recommendation describes how to measure the minimally required set of metrics to describe the performance of networks. Specification of advanced analysis and dissemination of measurement data are beyond the scope of this Recommendation.

Impairment allocation and mapping performance among IP and non-IP networks are beyond the scope of this Recommendation.

Methods to determine the exact network path that packets will follow are beyond the scope of this Recommendation.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T G.107] Recommendation ITU-T G.107 (2005), *The E-model, a computational model for use in transmission planning*.

[ITU-T O.211] Recommendation ITU-T O.211 (2006), *Test and measurement equipment to perform tests at the IP layer*.

[ITU-T P.800] Recommendation ITU-T P.800 (1996), *Methods for subjective determination of transmission quality*.

[ITU-T X.805] Recommendation ITU-T X.805 (2003), *Security architecture for systems providing end-to-end communications*.

[ITU-T Y.1540] Recommendation ITU-T Y.1540 (2007), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2006), *Network performance objectives for IP-based services*.

[ITU-T Y.1711] Recommendation ITU-T Y.1711 (2004), *Operation & Maintenance mechanism for MPLS networks*.

[ITU-T Y.1731] Recommendation ITU-T Y.1731 (2006), *OAM functions and mechanisms for Ethernet based networks*.

[IETF RFC 792] IETF RFC 792 (1981), *Internet Control Message Protocol*.
<http://www.rfc-editor.org/rfc/rfc792.txt>.

[IETF RFC 2330] IETF RFC 2330 (1988), *Framework for IP Performance Metrics*.
<http://www.rfc-editor.org/rfc/rfc2330.txt>.

[IETF RFC 2679] IETF RFC 2679 (1999), *A One-way Delay Metric for IPPM*.
<http://www.rfc-editor.org/rfc/rfc2679.txt>.

[IETF RFC 2680] IETF RFC 2680 (1999), *A One-way Packet Loss Metric for IPPM*.
<http://www.rfc-editor.org/rfc/rfc2680.txt>.

[IETF RFC 3357] IETF RFC 3357 (2002), *One-way Loss Pattern Sample Metrics*.
<http://www.rfc-editor.org/rfc/rfc3357.txt>.

[IETF RFC 3393] IETF RFC 3393 (2002), *IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)*.
<http://www.rfc-editor.org/rfc/rfc3393.txt>.

[IETF RFC 3432] IETF RFC 3432 (2002), *Network performance measurement with periodic streams*.
<http://www.rfc-editor.org/rfc/rfc3432.txt>.

[IETF RFC 3550] IETF RFC 3550 (2003), *RTP: A Transport Protocol for Real-Time Applications*.
<http://www.rfc-editor.org/rfc/rfc3550.txt>.

[IETF RFC 3917] IETF RFC 3917 (2004), *Requirements for IP Flow Information Export (IPFIX)*. <http://www.rfc-editor.org/rfc/rfc3917.txt>.

3 Terms and definitions

This Recommendation defines the following terms:

3.1 aggregate loss ratio: The loss aggregated along a path across multiple providers' networks.

3.2 demarcation point: Generally a point which separates two domains, here the separation between the access and transit networks.

3.3 fraction lost: The fraction of RTP data packets from source SSRC_n lost since the previous SR or RR packet was sent ([b-IETF RFC 1889]).

3.4 interarrival jitter: An estimate of the statistical variance of the RTP data packet interarrival time. The interarrival jitter J is defined to be the mean deviation (smoothed absolute value) of the difference D in packet spacing at the receiver compared to the sender for a pair of packets ([IETF RFC 3550]).

3.5 landmark system: A proxy system for customer premises terminal equipment.

3.6 measurement point: A point in the network containing functionality that may initiate or respond to measurements with other measurement points (located at peering points, demarcation points, PEs, CEs and landmark customer premises equipment).

3.7 path unavailability: The period of time from when losses exceed a threshold until they drop below another threshold.

3.8 period path unavailability: The total period of unavailability during a customer reporting period (typically 1 month).

3.9 probe: An individual IP packet associated with active performance testing, i.e., a test packet.

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

ALR	Aggregate Loss Ratio
ATM	Asynchronous Transfer Mode
BGP	Border Gateway Protocol
BSS	Business Support System
CE	Customer Edge
CoS	Class of Service
DoS	Denial of Service
DP	Demarcation Point
DSCP	DiffServ Code Point
DSL	Digital Subscriber Line
ECMP	Equal Cost Multi-Path
FEC	Forwarding Equivalence Class (multi-protocol label switching)
FSD	Flow Summary Data
GLONASS	Global Navigation Satellite System

GPS	Global Positioning System
IANA	Internet Assigned Numbers Authority
ICMP	Internet Control Message Protocol
IDQ	Inter-Domain Quality of Service
IPDV	Internet Protocol Packet Delay Variation
IPFIX	IP Flow Information eXport
IPLR	Internet Protocol Packet Loss Ratio
IPPM	Internet Protocol Packet Performance Metrics
IPPMS	Internet Protocol Performance Measurement Specification
IPSLBR	Internet Protocol Packet Severe Loss Block Ratio
IPTD	Internet Protocol Packet Transfer Delay
IPUA	Internet Protocol Unavailability
LAN	Local Area Network
LSP	Label Switched Path
MOS	Mean Opinion Score
MP	Measurement Point
MPLS	Multi-Protocol Label Switching
MSNG	Number of MiSsiNG probes
MTU	Maximum Transmission Unit
N_T_{max}	Number of probes exceeding T_{max}
NAT	Network Address Translator
NE	Network Element
NMS	Network Management System
NTP	Network Time Protocol
OAM	Operations, Administration and Maintenance
OSS	Operations Support System
PDV	Packet Delay Variation
PE	Provider Edge
PES	Performance Evaluation System
PL	Packet Lost
PLE	Threshold number of Probes Lost to End unavailability period
PLR	Packet Loss Ratio
PLS	Threshold number of Probes Lost to Start unavailability period
PoP	Point of Presence
PTP	Probe Transmission Period
PW	Policing Window
QoS	Quality of Service

RACF	Radio Access Control Function
RP	Rollup Period
RTCP	Real-time Transport Control Protocol
RTP	Real-time Transport Protocol
SDH	Synchronous Digital Hierarchy
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
SONET	Synchronous Optical Network
SP	Service Provider
SPW	Sliding Probe Window
TBD	To Be Determined
TCP	Transmission Control Protocol
TE	Terminal Endpoint
T_{max}	Maximum waiting time for a packet
TMF	Traffic Measurement Function
UDP	User Datagram Protocol
UTC	Coordinated Universal Time
VPN	Virtual Private Network

5 Performance attributes

It is important to recognize that the model of inter-domain QoS is an extension of the Internet architecture, which supports a connectionless IP service that delivers user payloads in the form of packets/bytes in each direction. Since outbound and inbound traffic routes may differ, the targets and measurements for all performance attributes in IDQ are one-way and reflect the connectionless nature of the service.

The performance attributes that are used to characterize the network performance (inter-domain QoS) of a path are:

- Mean one-way delay.
- One-way packet delay variation.
- Packet loss ratio.
- Path unavailability.

Each of these attributes has a corresponding performance parameter defined in [ITU-T Y.1540]. Although methods of measurement are beyond the scope of [ITU-T Y.1540], many relevant details may be obtained from the IETF IPPM RFCs (but noting that the IPPM metric definitions differ slightly in ways that do not detract from the sections on measurement procedures).

The list of attributes purposely omits some common metrics. For example, application throughput depends upon many factors including packet loss and transit delay, and others not under the control of the service provider (SP). Application throughput is not an independent IDQ performance attribute in its own right.

The offered traffic rate is also important as part of service descriptions and inter-SP contracts, but this is not considered a performance attribute.

Other performance metrics such as "delay equivalent to loss" and "packet reordering" are known to be useful. However, their incremental value over the metrics selected above is currently believed not to be worth the additional complexity they would require to specify, implement and deploy. Time may prove otherwise and other basic network metrics may be added in the future.

5.1 Mean one-way delay

Delay is important to the support of many applications including telephony, multimedia conferencing, financial transactions and online gaming. In addition, delay is indirectly related to throughput and impacts file transfer speeds and email delays.

The delay attributes of a QoS class are characterized by the mean one-way delay. Optionally, the minimum delay and a specific set of upper percentile delay variations may be provided. The percentile approach is used in preference to a standard deviation or variance model due to the occasional occurrence of bi-modal or multi-modal delay distributions.

The mean delay is defined as in [ITU-T Y.1540].

Delay is distance sensitive due to the non-infinite signal propagation speed. Mean delay may vary between QoS classes due to priority queuing which is taken into account when setting objectives.

5.1.1 Relationship to existing standards

The mean delay specified here would correspond to "Type-P-Finite-One-way-Delay" if extrapolated from [IETF RFC 2679]. This metric requires a conditional distribution of delay (conditioned on arrival within a fixed waiting time that is set long enough to distinguish packets with long delays from those that are truly lost, e.g., discarded or corrupted), where lost packets have undefined delay.

5.2 One-way packet delay variation

In addition to the mean delay, packet delay variation (PDV) is important to many applications including telephony, gaming and transactions.

Packet delay variation is defined in [ITU-T Y.1540], and is essentially the difference between the 99.9th percentile of delay and minimum delay. The minimum delay of a sample or set of individual delay measurements is taken as the reference for variation. The percentiles of the actual delay distribution can be estimated by summing the minimum delay and the delay variation percentile of interest.

Other useful delay variation percentiles that can be recorded include:

- 90th percentile – DV90.
- 99th percentile – DV99.

Conceptually, percentiles are measured by stack-ranking all measurements of successfully delivered packets, discarding a top percentage e.g., 0.1% in the case of 99.9th percentile, then selecting the remaining highest value. In reality, we measure a subset of packets, the active probes. All lost packets or packets delivered while the network is considered unavailable are ignored from other metrics.

By taking multiple percentile readings and a minimum delay reading, the distribution of delays can be better understood. This information is more useful than a simple standard deviation metric which can be easily used only when assuming a mathematically friendly underlying probability distribution function. In reality, network delay characteristics may be multi-modal. There is added complexity and cost associated with engineering a network to closely match a particular delay distribution and in measuring that distribution accurately. Therefore only QoS classes that require multiple percentiles will have them specified and measured.

Delay variation is loosely correlated to distance (since distance is loosely correlated to number of hops) and allows targets to be set independently of site locations. Delay variation is correlated to link bandwidth and utilization and therefore, access links are a common source of delay variation, owing to their low bandwidth with respect to other network links.

5.2.1 Relationship to existing standards

The delay variation specified here is based on observations of the delay distribution, rather than on the difference between two successive delay measurements as is the usual formulation associated with [IETF RFC 3393]. However, [IETF RFC 3393] has sufficient flexibility to produce either the inter-packet delay variation or the delay variation by using a fixed minimum delay reference.

5.3 Packet loss ratio

Packet loss is important to most applications. It significantly impacts either quality or throughput of the network.

A packet is considered lost if conditions satisfy the qualifications of a lost packet outcome as defined in [ITU-T Y.1540]. The maximum waiting time threshold, T_{max}, should be set long enough to differentiate a packet with long delay from a packet that is truly lost, e.g., discarded or corrupted. In practice, this waiting time may need to be set at 10 seconds or more, and requires knowledge of the network under test to set correctly.

IP packet loss ratio is the ratio of total lost IP packet outcomes to total transmitted IP packets in a population of interest, as defined in [ITU-T Y.1540]. The population of interest is usually the total set of packets sent between a particular source and destination with specific type, payload size, DSCP and class assignment, etc.

Packet loss is largely insensitive to distance and objectives can be set independently of the end site locations. Packet loss ratios are sensitive to access technologies, bandwidth and utilizations, and number of hops, and objectives must be set accordingly.

5.3.1 Relationship to existing standards

Y.1540 packet loss differs from Type-P-One-way-Packet-Loss as defined by the IETF IPPM, in that errored packets are designated lost in [IETF RFC 2680]. In practice, this difference is not significant to measurement results because packets with errors are usually discarded before they reach the destination. However, if the last link entering the measurement point is error prone, then the difference between the [IETF RFC 2680] and [ITU-T Y.1540] definitions may be significant.

5.4 Path unavailability

Path unavailability is significant when a human observer detects a business-impacting application failure due to network loss. For a typical application such as telephony, a network path is considered unavailable by the user if there is an inability to connect, or a connection is lost. The measurement of unavailability attempts to approximate this view by detecting periods during which network path unavailability would have noticeable impacts on applications and individual or business productivity.

Unlike delay and loss attributes, the unavailability attribute is not statistically simple to define and an approximation is required.

In IDQ, unavailability is calculated from the distribution of loss measurements over time; see clause 7 of [ITU-T Y.1540] for details on the service availability function. A period is considered unavailable if there is an excessive packet loss ratio (PLR) (e.g., >75%) over a specific fixed time interval. The interval may be set independently for each QoS class, but currently only one interval is specified, 5 minutes.

This definition is intended to capture periods of very poor performance and requires the network performance to return to normal levels before the unavailability is ended. During a period of unavailability, none of the delay or delay variation metrics are valid.

Path period unavailability (IPUA) is measured by summing the periods of unavailability and dividing by the total period being covered. The period being covered to be used is the default "reporting to customer" period. It should be noted that the IDQ system keeps track of each individual period of unavailability for reporting to customers.

Unavailability is largely insensitive to distance, but is sensitive to single points of failure in a network architecture. It will vary significantly with access technologies and configurations. To achieve a low level of unavailability, diverse transmission paths are required.

5.4.1 Relationship to existing standards

[b-IETF RFC 2678] defines parameters for unidirectional connectivity and bidirectional instantaneous connectivity. Both these metrics can be used to assess connectivity over time, similar to the Y.1540 service availability function.

6 Performance measurement requirements

Inter-domain QoS is intended to increase the level of confidence in the expected service characteristics of the NGN. Increased confidence will enable new applications, services and revenue streams. An integral part of achieving this confidence is the continuous measurement of service performance. The purpose of taking measurements is to provide information for customers, potential customers and service providers, and includes:

1) For customers and potential customers:
   a) Reports to customers of what service has been delivered.
   b) Reports to potential customers to support marketing claims on service characteristics.
2) For service providers and third party delivery assurance entities:
   a) Reports to design service offerings.
   b) Reports for troubleshooting.
   c) Data for marketing collateral.
   d) Reports to enable capacity planning and service development.

The IDQ measurement system and the statistics that it produces must:

a) be easily understood by SPs and customers;

b) be well defined (non-ambiguous);

c) be relevant to customers' applications;

d) enable service providers to diagnose issues and anticipate capacity requirements;

e) be independently repeatable (multiple SP measurements over the same time get the same result);

f) be independently verifiable by customers (customer measurements should be close to SP estimates);

g) be widely applicable (traffic type, link size, load independent, any IP network);

h) be appropriately sensitive to distance and path;

i) not significantly impact the forwarding of other data;

j) be sufficiently scalable to support millions of customer sites;

k) be sufficiently reliable to enable SLAs with financial penalties to be administered;

l) be sufficiently accurate to enable SLAs with financial penalties to be administered.

Since outbound and inbound traffic routes may differ, all measurements will be "one-way". Customers or service providers may aggregate the statistics of two directions to estimate the round-trip performance.

Measurements will be taken from each of the segments of the measurement network model (described in clause 8) and may be combined to form multi-segment, site-to-site, edge-to-edge or IPTerminal-to-IPTerminal metrics. A subset of these metrics will be used for reports for the offered services.

Quantitative requirements for end-to-end and segment accuracy have not yet been developed. The following incomplete list of measurement aspects should be considered when requirements are set, and when systems and components are designed:

- Number of segments (due to concatenation errors).
- Impact of measurement equipment not being directly in user data path.
- Measurement equipment processor load.
- Time synchronization errors.
- ECMP-related errors.
- Measurement granularity (unit).
- Number of measurement samples per evaluation period to support required statistical accuracy.
- Active probe frequency.
- Active probe size.

6.1 Active measurement requirements

The performance of active probes will be used as a predictor of the performance of users' data. Time-stamped delay and loss measurements will be collected. Probes will be injected into the network at certain devices and sent to extracting devices which will return the measured information to the injection device.

6.1.1 High level requirements

The probes will be:

a) UDP-echo based.

b) Usable for the measurement of both delay and loss, preferably in both directions between two devices.

c) Marked with the appropriate DiffServ QoS class, preferably both in the header and body for each direction.

d) Preferably transmitted at periodic intervals with pseudo-random start times near the beginning of the evaluation interval.

e) Time-stamped at injection and extraction devices.

f) Preferably marked with source and destination addresses from address pools (to minimize impact of load-balancing).

g) Able to indicate a loss in confidence of local clock sync back to initiating device.

h) Probe packets should be able to be marked with the appropriate MPLS EXP bits if the underlying network uses MPLS technology.

6.1.2 Specific requirements

[IETF RFC 3432] requires that a periodic sequence is started with a small random variation from the specified start time and subsequent probes each keep the same offset from UTC.

A separate set of probes will be used for each of the IDQ network QoS classes. Packet size is selected to represent the users' packets in each QoS class. The current recommendation is as follows:

Table 1 – Probe packet size for selected network QoS Classes

Network QoS class	Description	Probe payload size (octets)
Class 0	Telephony	20
Class 2	Low latency data	256

Probe packet sizes for other QoS classes are given in Appendix I.

NOTE – The probe size should be constant when measuring delay variation, since the delay metric includes probe serialization time, and serialization time will vary with probe size possibly causing error in the assessment of delay variation.

Consideration of the pattern of inter-probe timing is important. The current recommendation is to use continuous probing with equal inter-probe interval referenced to the first bit of each packet.

The following segment metrics are derived from the probe delay, probe loss and probe timestamp measurements:

a) Mean delay.

b) Minimum delay.

c) 99.9 percentile of delay variation (90, 99 percentiles optional).

d) Unavailability.

e) Loss ratio.

The inter-probe transmission period is determined by the number of measurement samples required for sufficient accuracy of delay percentiles. This will be referred to as the probe transmission period (PTP). The PTP may be different for each QoS class and by default is 100 ms. Measurement samples are aggregated over a period of time to be referred to as the rollup period (RP). The rollup period for all measurements will be 5 minutes.

The start of RP is synchronized among all participating SPs to coordinated universal time (UTC) and is based on the beginning of each UTC hour. Accuracy is derived from the global positioning system (GPS).

An estimated average probe rate of 1000 probes per 5-minute rollup period is to be used for all percentiles and QoS classes. This includes an allowance of 1% for lost probes. The estimated probe rate will be validated before deployment since too low a choice impacts accuracy and too high a choice wastes resources.

(Note that with this probe packet rate, the minimum loss ratio that can be reported is 10^{-3}, and this may not be sufficient to characterize some QoS classes accurately, such as Y.1541 classes 6 and 7. More study is required for measurements in those classes.)

Looking at the bandwidth consumption that each-way probing consumes, assume:

- An average of 10 probe packets per second.
- Measurements of 3 network QoS classes.
- Using 64-byte probe packets.

Each probe stream consumes 5,120 bit/s, so for 3 QoS classes the total probe stream is 15,360 bit/s. This is 0.003% of the total traffic of an OC-12/STM-4 link, 1% of a T1 link or 0.8% of an E1 link.

A typical CE having IDQ service would use two-way probing. Total probe stream traffic on the CE-PE link would be 15,360 bit/s in each direction.

The bandwidth consumption within a backbone is dependent upon the number of probe streams. The purposes of different probe streams are described in clause 7.2. Once a probing scheme has been designed, the evaluation of bandwidth consumption may occur.

6.1.3 OAM-based active measurement requirements

This clause provides several examples of possible OAM packet formats which meet the requirements for OAM-based measurements. The examples are shown in this clause strictly to clarify the usage of OAM for passive measurement. The standardization of the OAM packet formats and their semantics are beyond the scope of this Recommendation.

- Example of MPLS OAM packet format (see [ITU-T Y.1711]).
- Ethernet OAM packet format (see [ITU-T Y.1731]).
- ICMP packet format (see [IETF RFC 792]).

6.2 Passive measurement requirements

6.2.1 High level requirements

The passive measurement general requirements are:

- Passive measurement entity shall be one of the following: network element resident measurement entity or standalone measurement entity.
- Every single measurement shall have at least source and destination addresses, an associated QoS metric, and accurate starting and ending time.
- Timestamps should be traceable to UTC and sufficiently accurate to meet the requirements in clause 6.6.
- Passive measurement shall capture a copy of the traffic without introducing modifications in the original traffic.
- Passive measurement shall classify traffic in different granularity (e.g., 5-tuple, VPNID, IPv4/6, etc.).
- Passive measurement should support probabilistic (e.g., random) and hash sampling methods.
- Passive measurement should support flow-based sampling methods.
- Passive measurement shall perform sampling operation at wire-speed.
- Passive measurement should support sampling both before and after classification.
- Passive measurement should measure the performance of fragmented packets.
- Passive measurement shall have the capability to measure various packet sizes, up to the maximum MTU for a path.
- Passive measurement shall derive various performance metrics such as delay, jitter, packet loss and unavailability.

6.3 Measurement time-escales

We now consider a common option, namely time-escales. Inter-domain QoS requires that all performance metrics are measured over the same time-escales. This greatly simplifies analysis of inter-domain performance.

The selected time-escales for performance measurement support the following criteria:

- The overhead load due to measurement traffic must be kept at a low level.
- The basic time-escale must be large enough to contain the start and end of a large number of traffic flows.
- The basic time-escale must be common and synchronized globally among SPs (preferably independent of network characteristics which may impact the timing path, e.g., link/network element failures, congestion and delay variation).
- The time-escale must be meaningful to network users and capture any productivity or service quality issues they perceive in the network.
- The time-escales should not unduly emphasize momentary glitches such as link outages or rerouting events where they do not significantly impact network user experience.

Given these criteria, the default time-escales selected are:

- Measurement: Time-escale unit is 5 minutes. This is synchronized via GPS or similar service, and aligned with UTC. This allows all SPs to synchronize their measurement periods and correlate measurements. The targets and measurements for mean delay, delay variation and packet loss apply to 5-minute periods. Measurement samples are aggregated over this period of time which is referred to as the rollup period.
- Customer reporting: Time-escale unit is 1 "month" with start and end hour/day defined by the SP offering the IDQ service. The start and end monthly definitions may not be aligned between SPs. To be able to correlate measurements from one time zone to another and one SP's "month" to another, all timestamps are referenced to UTC as well as any local time references. The actual time-escale of customer reporting needs to be determined by agreement between the network provider and each customer.

6.3.1 Relationship to existing standards

[IETF RFC 3432] refers to the rollup time as defined above as "Tcons", a time interval for consolidating parameters collected at the measurement points.

[ITU-T Y.1541] refers to the "rollup period" as the "evaluation interval" and suggests an evaluation time of 1 minute for IPTD, IPDV and IPLR.

6.4 Measurement system unavailability

If parts of the measurement system itself are unavailable, that will inhibit the ability of the provider to demonstrate that his QoS targets have been met during the period of unavailability. However, it is almost certainly not as serious for the measurement system to be unavailable as for the IDQ service itself to be unavailable as defined above. We therefore suggest that while unavailability of the measurement system should be tracked, it should not be automatically treated as equivalent to unavailability of the service. In the event that a customer claims that an SLA target was violated during some measurement interval, the provider would normally have measurement data to show how his segment of the network was performing at that time. If the provider cannot produce data to show that SLA targets were being met because his measurement system was not operational during that interval, he may have no choice but to assume that he did in fact violate the SLA. Thus providers will be highly motivated to keep their measurement systems operational all the time but will not automatically be penalized for measurement system outages.

6.5 Interaction of policing and performance measurement

Ingress and egress segment performance is sensitive to the level of customer traffic. The performance levels of each IDQ network QoS class can only be delivered assuming that the traffic is within the subscription bounds for that QoS class.

In the event that traffic does exceed its subscription bounds, packets may be delayed, discarded or have their DSCP remarked. These actions will potentially change the delay and loss characteristics of the data streams as well as any UDP echo probes that traverse the policing point. There is no fail-safe mechanism to detect which UDP echo probes are impacted by a policing event.

To handle the interaction between policing and performance measurement, inter-domain QoS discounts measurements taken during a period when there is a policing-detected violation for that QoS class.

The determination of when policing-detected violations occur for a network QoS class is made through SNMP polling of the DIFFSERV-MIB. The DIFFSERV-MIB keeps a counter of any policing-detected violation in each QoS class and, by comparing the counters at the start and end of the policing window (PW), the determination is made whether any policing-detected violations occurred. The interaction/communication between routing systems that perform policing and network measurement systems (that must indicate when measurements may be affected by the policing operation) should be accomplished in the measurement management system (see Appendix II).

The policing window (PW) is the periodic rate at which SNMP polling takes place and by default is 5 minutes for each QoS class.

If a policing-detected violation occurs for a QoS class during a policing window, the delay, loss and availability statistics for that rollup period shall be marked as possibly affected by an excess traffic condition. Measurements that are collected during customer traffic overload may not be suitable for comparison with service level agreements (SLAs), but might be useful for other purposes. The list of these rollup periods, and the associated number of packets that exceeded the agreement for each QoS class, is kept. These details and an aggregate of the total time and total exceeding packet count are reported to customers.

This method encourages customers to subscribe to the appropriate level of bandwidth in order to ensure that their QoS class characteristics are maintained at all times.

Policing-detected violations between SPs will similarly be detected and reported each rollup period.

6.6 Clock synchronization

Clock synchronization specifies the extent to which multiple clocks agree on the time.

It impacts:

1) common understanding of when a measurement or event occurred or is planned to occur;

2) accuracy of certain network performance measurements.

The magnitude of time offset between measurement points is critical to the accuracy of the one-way measurement attributes of minimum delay, mean delay and delay percentile. The attributes of delay variation and loss are unaffected by offset magnitude. Unavailability is unaffected although there may be minor inaccuracies in the reported time of occurrence.

The measurement points per the network models can be grouped into three categories:

1) Demarcation and peering measurement points.

2) CE and PE measurement points.

3) Customer host measurement points.

We allocate a maximum offset to each category:

1) The clock of demarcation and peering measurement points can have an offset from GPS of no more than 100 ms magnitude.

2) The clock of the CE router, PE router or co-located non-router measurement device can have an offset from its paired demarcation measurement point of no more than 1 ms magnitude.

3) The clock of certain customer host measurement points can have an offset from its paired CE router of no more than 1 ms magnitude.

If a PE is also a demarcation point then the tighter offset is to be applied.

Providing clock synchronization at these points supports the measurements described in clause 8.

Figure 1 below shows these three categories and where the specified offsets apply.

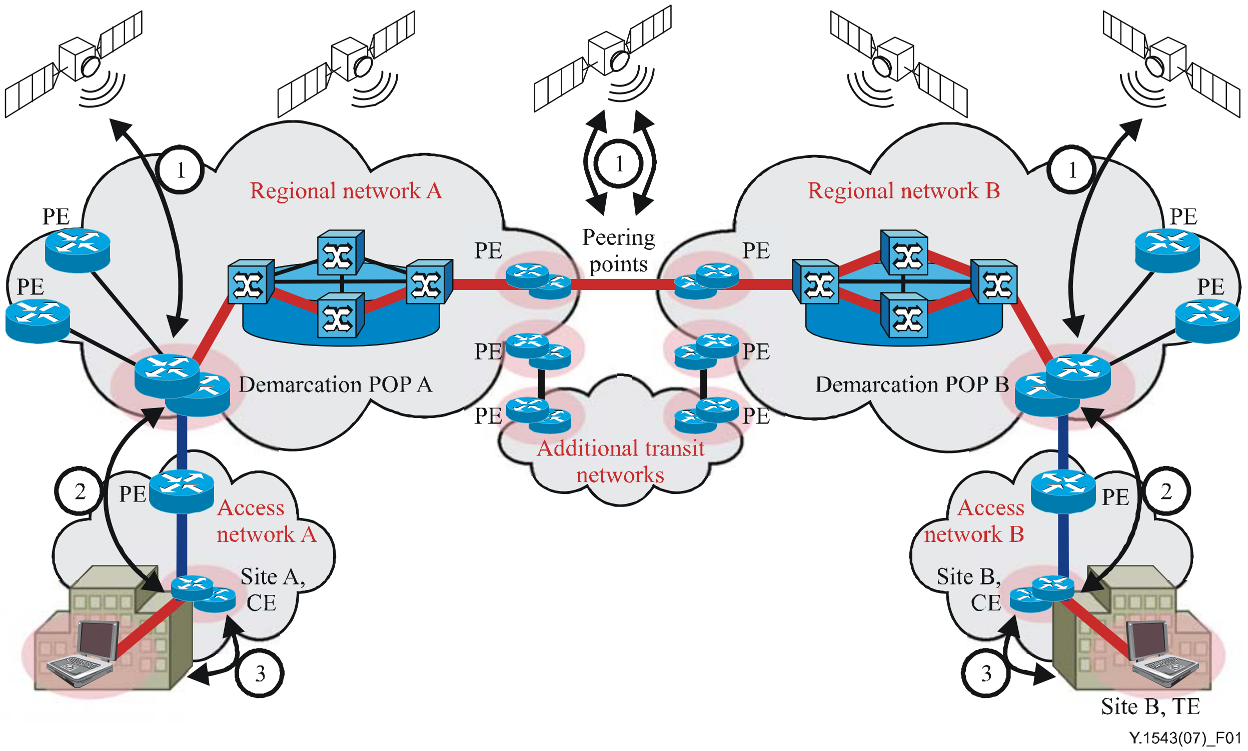


Figure 1 – Three categories of equipment at which offset maximums apply

6.6.1 Relationship to existing standards

[IETF RFC 2330] describes clock terminology and wire time.

We use the term "synchronization" per [IETF RFC 2679]. Synchronization measures the extent to which two clocks agree on what time it is. [IETF RFC 2679] loosely maps the IPPM group's terminology to ITU-T's terminology (e.g., [b-ITU-T G.810] and [b-ITU-T I.356]). It analyses measurement errors.

[IETF RFC 3393] discusses the minimal impact of clock synchronization on differential measurements, of which delay variation is an example.

6.6.2 Implementation methods

GPS is used as a reference; however, other implementation methods (e.g., Galileo, GLONASS) may be used to synchronize demarcation and peering points as long as the offset to GPS requirements are met. In cases of inadequate reception, the use of pseudolites or other techniques to provide accurate clocks derived from GPS may be required.

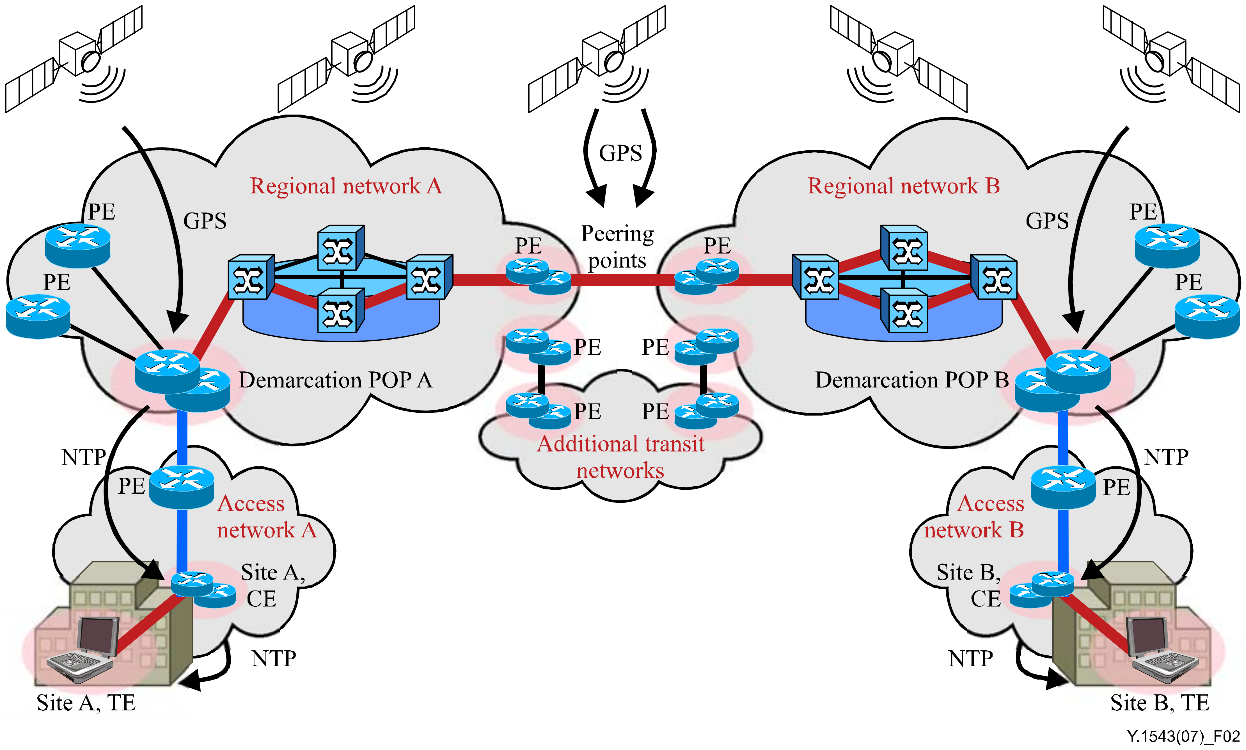


Figure 2 – Example of clock synchronization implementation

Figure 2 shows an example clock synchronization configuration for a TE-TE scenario, where:

1) GPS receivers are used to set the time of shadow routers at both demarcation POPs.

2) NTP is used to set the time at the CEs, which are NTP clients using a demarcation POP's shadow router as an NTP server.

3) NTP is used to set the time at selected customer hosts, which are NTP clients using the CE as an NTP server.

CE clock synchronization should be via NTP to the SP's closest GPS system as client. This may not be their associated demarcation POP. Since NTP offset from client to server is a function of delay asymmetry between client and server, using NTP in some cases may not meet the clock synchronization offset requirements, in which case alternatives must be found.

CE routers may be used to provide a multi-homed IDQ service from single or multiple SPs. In any case, clock synchronization should be set using "prefer" via NTP from the closest measurement POP. It will automatically switch upon loss of synchronization.

This Recommendation does not address how SPs may set up GPS and NTP to meet these requirements, nor how to validate the offsets of their systems relative to a GPS-derived time. These topics will be the subjects of other Recommendations.

6.6.3 Loss of synchronization

Measurement points should be able to detect when they have low confidence of being adequately synchronized, e.g., if the NTP server becomes unreachable, measurement points should:

1) notify a management station;

2) provide other measurement points with information about the probes to which they are responding.

There are several degrees of timing inaccuracy in the context of performance measurements. There is always some error between the clocks of the sending and receiving measurement device:

- Some degree of error is tolerable for all measurement types (e.g., < 1 ms).
- Some degree of error is tolerable for loss measurements, but excessive inaccuracy for one-way delay measurements would make them useless (e.g., > 100 ms).
- Some degree of error causes failure for all measurements (e.g., 1 s).

Loss measurements have a large tolerance to time error because of the substantial "waiting time" for packets to arrive at the destination. While normal one-way delays are < 400 ms, the waiting time to declare a lost packet is usually 3 seconds. Therefore, loss measurements are less susceptible to errors in time synchronization when compared to delay measurements. It is important to detect synchronization issues (low confidence) and record this condition with all the results, including the degree of error if it can be determined.

6.7 Measurement granularity

QoS in NGN can be provided at various levels depending on service requirements. Its granularity can be as fine as a flow-level or as coarse as CoS (class of service) level. More specifically, the granularity levels consist of a flow, various layer 2 tunnels (e.g., L2TP, L2VPN, etc.), an MPLS LSP, layer 3 tunnels (e.g., GRE, IPsec, L3VPN, etc.), any other class-based logical paths (e.g., an IP path associated with DiffServ class), and an application session. Various mappings are possible among them. For example, a number of flows can be aggregated to form a tunnel. Several tunnels or logical paths may represent an application session.

There are several definitions of the term "flow" being used. This Recommendation adopts the definition used in [IETF RFC 3917] as follows:

A flow is defined as a set of IP packets passing an observation point in the network during a certain time interval. All packets belonging to a particular flow have a set of common properties. Each property is defined as the result of applying a function to the values of:

1) one or more packet header field (e.g., destination IP address), transport header field (e.g., destination port number), or application header field (e.g., RTP header fields [IETF RFC 3550]);
2) one or more characteristics of the packet itself (e.g., number of MPLS labels, etc.);
3) one or more of fields derived from packet treatment (e.g., next hop IP address, the output interface, etc.).

A packet is defined to belong to a flow if it completely satisfies all the defined properties of the flow.

Flow-level performance measurement may be needed for a high quality user service which needs special care or has a special billing purpose. Due to the performance complexity, it may not be practical to have continuous real-time flow-level measurement for all the service flows. However, it is necessary to define such functional capabilities to meet a special service requirement. Fortunately, it may be possible to meet both flow level measurement and scalability requirements. If we measure entire flows at a particular measurement point of interest, it is not scalable. Typically, meaningful flows which take most traffic volume (e.g., over 95% of a particular link bandwidth) comprise of small portion of the entire number of flows. Thus, if we can identify these meaningful flows, we can measure them in real-time continuously and avoid measuring unnecessary flows. Measurement on tunnel level and other higher level paths introduce much less stringent performance burdens and thus scalability is not an issue.

Tunnel level, logical-path level, as well as application-session level measurement also needs to be supported to meet various measurement requirements such as tunnel statistics, per-class statistics, and per-application statistics. These measurements are meaningful for an entire end-to-end path whether it is a tunnel, a logical path, or an application session. Flow, tunnel and logical path level measurement is relatively clear on how to measure them since each one has a unique means of identification. However, an application-session level measurement requires mapping or aggregation of lower-level measurement results. For instance, a video telephony session can be composed of voice and a video flow. Each flow can be a class-based logical path or part of a L2VPN path.

The selected granularity for performance measurement shall support the following criteria:

- The measurement overhead must be kept at as low a level as possible.
- The measurement may support all levels of granularity as described above.
- Both active and passive measurement methods can be used as applicable.
- The flow-level and other fine-grained (e.g., LSP-level) measurement shall be supported on a demand basis.
- The flow-level measurement should have an end-to-end context. Concatenation of segment-based flow measurement may not reflect the original flow characteristics.
- The measurement may support relevant levels of granularity for multicast traffic.

7 Measurement network model

Ideally, measurements to assure performance of customer traffic would be taken between the same endpoints as each customer's traffic. Whether these endpoints are the customer's terminal (TE), customer edge router (CE) or provider edge router (PE), the number of measurements would be so great as to make this impracticable. Therefore, we look to a practical solution and find one by segmenting the network into a measurement network model.

Segmenting a network is a trade-off between the following requirements:

- Minimize cost.
- Support service flexibility.
- Ensure accurate end-to-end measurements.
- Support measurement comparison to each providers' impairment target.

Costs associated with each segment include (assuming one-way active probing):

- Clock synchronization at each segment end.
- Initiation and response of probes at respective segment ends.
- Associated measurement data which needs retrieval, storage and distribution.
- Contribution to concatenation error.

The greater the leverage of a single measurement produced by a segment probe, the fewer probes will be needed. If fewer segment measurements may be used in the calculations of thousands of concatenated estimates, then there will be lower total probe overhead.

Providers offer assured delivery services between different endpoints. We use the shorthand terminology.

1) "edge-edge" for services that extend to the edge of a providers' network.
2) "site-site" for services that extend to the edge of a customer's premises (this is sometimes called end-to-end).
3) "TE-TE" for a managed customer network service, we will consider this as extending to a customer's terminal. We note that some service architectures place one instance of TE within the traditional network boundaries, e.g., IP television services.

All three services must be supported by the models. There is no requirement that both endpoints have similar services (i.e., demarcation points). This terminology is used to emphasize the distinction in endpoints. Network segmentation provides service differentiation opportunities to providers who may offer assured delivery and reporting for a subset of segments.

The models must support measurements which will enable comparison of measured performance to impairment targets. Measurement points located at CE or PE locations may use capabilities of the CE or PE routers themselves or separate co-located measurement equipment.

We note that NGN terminology differs from the communication industry's terminology used in this Recommendation, and so provide the table below with mapping between the terms.

PE	one of	access node, access border gateway, edge node, or interconnection border gateway
CE	maps to	(new term) customer premises edge node

7.1 Network partitioning

The network is partitioned into segments, each being monitored independently. This partitioning enables the scaling of the network with sub-linear growth in the amount of monitoring traffic and equipment relative to the number of customer sites involved.

Typically, the network is considered to consist of ingress and egress access segments, and a transit segment. It is assumed that one regional service provider will provide an access network that supports both ingress and egress segments for a specific site. There may be a backbone service provider(s) providing transit services between the regional service providers.

A specific service provider may act as either or both an access provider for some traffic and as a transit provider for some traffic. A demarcation point between access and transit segments is named a "demarcation POP" (DP).

Demarcation points at the customer end of the ingress and egress segments are dependent upon the service.

- For "edge-edge" services, demarcation points are typically PEs.
- For "site-site" services, demarcation points are typically managed CEs.
- For "TE-TE" services, demarcation points are typically customer's terminals.

These demarcation points are illustrated in the following Figures 3, 4 and 5, where the models are named "edge-edge", "site-site" and "TE-TE".

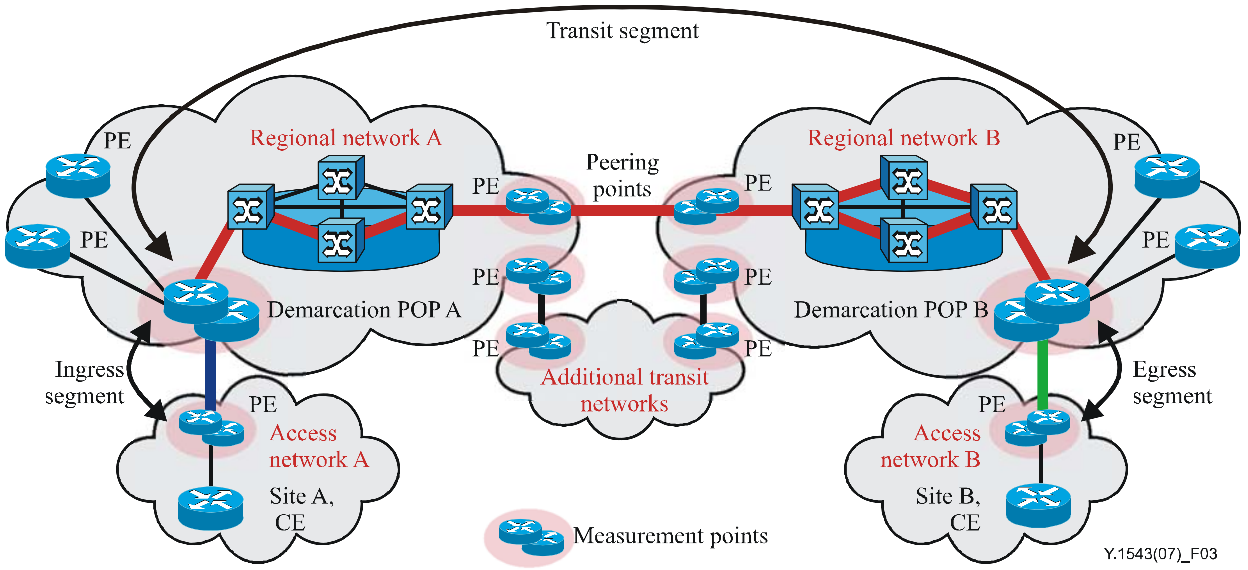


Figure 3 – Edge-edge model

In the edge-edge model, delivery is assured to a PE nearest a customer, service between customer terminals or CE to the PE is not assured. The assured performance characteristics of the network are comprised of the aggregate of the performance characteristics of the ingress, transit and egress segments.

The ingress and egress segments do not include the CE-PE link, but do include the provider edge router as well as regional switching and transport.

The transit segment is measured from demarcation POP of the ingress regional service provider to the demarcation POP of the egress regional service provider. This segment may or may not include separate backbone service providers. The transit segment may span a city, country/state, continent or multiple continents.

The transit segment may include parts of the ingress and egress regional networks, interconnects between the regional and backbone providers, and transit service across any backbone networks. The transit service of the backbone network is a sub-segment of the entire transit service.

Partitioning of measurement responsibility may follow network boundaries. However, measurement responsibilities may cross boundaries in any configuration that serves the goal of complete measurement coverage. For example, two or more networks may be covered by a single measurement system's measurement points.

The models support multiple peering connections between providers. Only one is shown for simplicity. The models support equal cost multipath (ECMP) as indicated by the multiple paths shown within providers. In many instances, there may be multiple paths over which traffic may traverse. By having probes follow a plurality of paths, performance contributions from each path will be included in the reported statistics. Covering this path diversity as part of the measurement is achieved by using a range of addresses for each demarcation POP. Each of which will be configured to respond to probes sent to any of 16 addresses and will be able to send probes sourced from any of 16 addresses. This will support a total of 256 flows, which increases the likelihood that, in the case of load balancing, active probes will follow all the paths that a customer's data follows between 2 sites.

Note that results collected using multiple addresses cannot be pooled for metrics such as delay variation; otherwise the results would not be representative of customer flow performance.

Since there is limited load balancing expected between CE or PE and the demarcation POP, the CE/PE need only have one address, which in combination with the 16 addresses of the DP's measurement device will provide sufficient route diversity to include measurement contributions from all load balanced paths. If the CE/PE is configured to probe across the transit segment then 16 addresses would be preferable.

This approach to ECMP emphasizes coverage of all the paths that can be seen. Other approaches conduct measurements on a subset of paths which are representative of users' traffic.

The ingress, transit and egress segments are monitored from demarcation POPs that are specifically located for the role. Demarcation POP selection is an SP choice. Each customer site is assigned to a demarcation POP within its regional providers' network. The POP is selected on the basis that the majority of the traffic from that site to others goes through this specific POP, which is within the same geographical region as the customer site. There is a minimum number based on the location of customer sites. SPs may increase the number of measurement POPs as they see fit, and some SPs may elect to make every PE POP a measurement POP.

The demarcation POP will have one or more measurement systems. It will monitor the backbone network and initiate tests with PE and CE devices. Thus it will be capable of measuring ingress, egress and transit segment performance. It will also collect and collate all necessary statistics.

Inter-domain QoS relies on the ability to collect inter-service provider statistics on a continuous basis and for service providers to be able to resolve the causes of performance targets not being met. To support this monitoring and troubleshooting requirement, there are a set of requirements that must be met by service providers:

- Each participating provider must provide measurement points that act as performance characteristic test points for their use, and possibly for restricted use by other SPs.
- Measurement points must be located at any participating service providers' major interconnection peering POP.
- There must be a measurement point (demarcation POP) nominated by regional providers for each participating customer site.
- A service-dependent measurement point at PE, and possibly at CE and/or customer TE if this scope of performance assurance is supported.

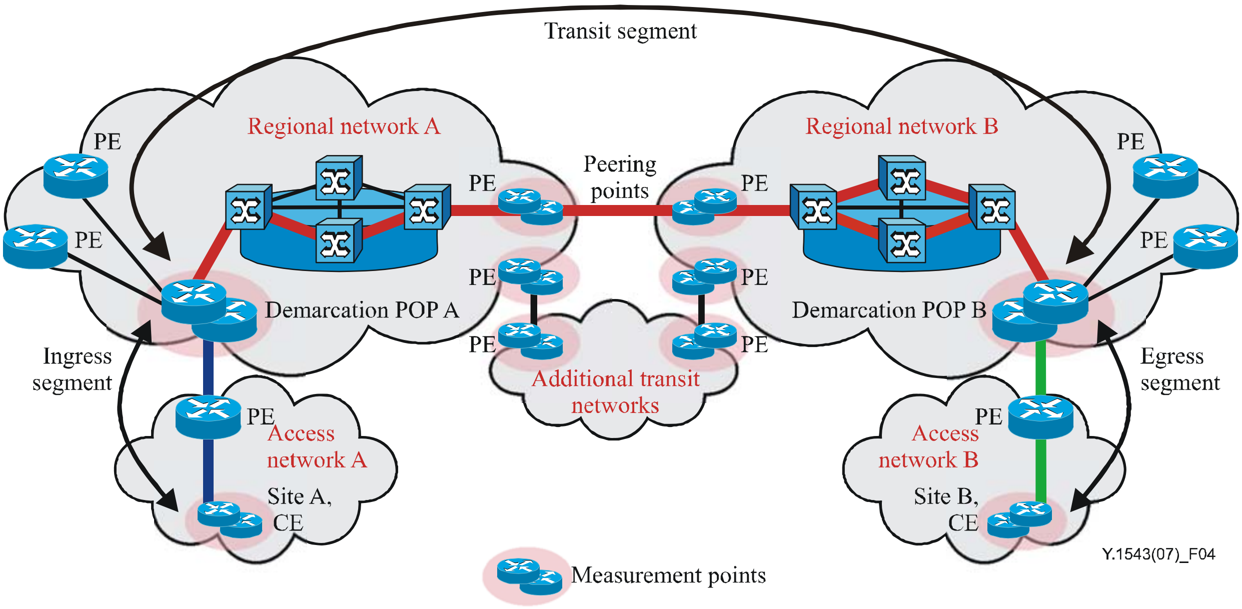


Figure 4 – Site-site model

In the site-site model, delivery is assured to customer CE, service between customer terminals to the CE is not assured by the service provider. It is the responsibility of the customer. The assured performance characteristics of the network are comprised of the aggregate of the performance characteristics of the ingress, transit and egress segments.

The ingress and egress segments include an access segment (DSL, cable, SONET/SDH, Ethernet, etc.) including the customers edge (CE) router as well as regional switching and transport.

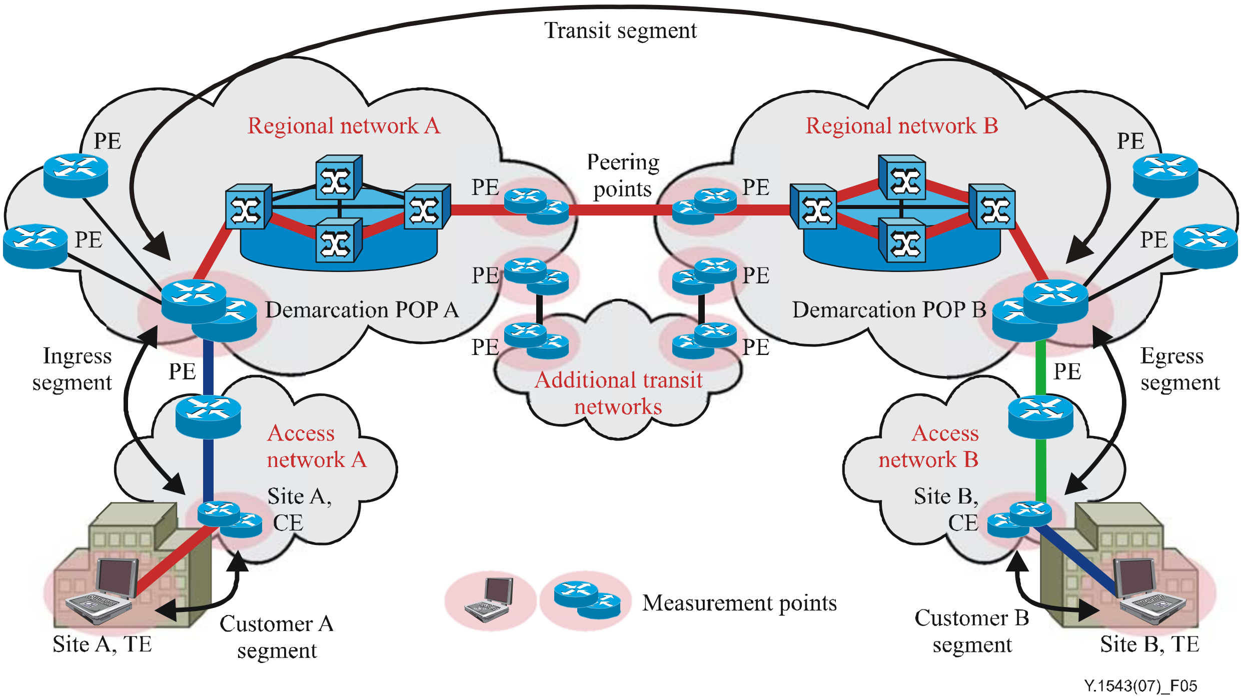


Figure 5 – TE-TE model

In the TE-TE model, the assured performance characteristics of the network are comprised of the aggregate of the performance characteristics of the ingress, transit egress and customer segments.

The customer segment includes the network between a CE and a customer's TE. This may include home networking arrangements to company LANs, computers and appliances.

Selection of the customer's TEs to be used for measurements include consideration of:

1) Stability:
   a) static address or directory lookup;
   b) stationary rather than mobile;
   c) always online.
2) Performance:
   a) probe response not impacted by other programs.
3) Clock synchronization:
   a) required for one-way delay and delay percentile measurements.
4) Representativeness of many other TEs:
   a) analysis or measurement may show that measurements between a CE and a particular TE is representative of many other TEs, called "landmark" TEs.

5) Number of TEs probed:

   a) to minimize the number of probes, a minimum number of landmark TEs should be used;

   b) to minimize the complexity of data handling and reporting, a minimum number of landmark TEs should be used.

Communication from a CE to a TE may require NAT traversal. Depending upon the administration of these devices, pre-provisioning or NAT traversal protocols may need to be used. Alternatively, the NAT device may be used as a measurement point as a proxy for TEs.

It is expected that there will be cases when there will be very little performance variation in the customer's network. In these cases, instead of the use of operating measurements, fixed impairment values may be agreed to.

7.2 Applied measurements

Measurement purposes fall into three broad categories, operating, supporting and testing.

- **Operating** measurements are those which are made on an ongoing basis between measurement points to monitor normal operation of the assured segments along customers' data paths, e.g., measurements of ingress, transit and egress segments.
- **Supporting** measurements, which may be taken continuously, are used to provide information for SPs. These measurements occur in addition to operating measurements and can be between various measurement points, e.g., measurements of each SPs' contribution to the transit segment.
- **Testing** measurements are made on an exception basis following the detection of abnormal operating measurements for troubleshooting or to test a new path. These measurements occur in addition to operating or supporting measurements and are between measurement points which do not have operating or supporting measurements being taken, e.g., measurement of a particular CE-to-CE path for a prospective customer.

Some measurements may fall into multiple classes. For example, a CE-to-CE measurement may be used for a prospective customer (testing), as a sanity check for providers (supporting), or as a premium (un-scalable) customer service (operating).

Different views of the same measurement data may be useful for different purposes. For example, a provider that collects and analyses ongoing measurements at sub-intervals of RP may evaluate the impact of remedial action upon network performance more quickly than had they waited for the RP before doing so.

The following scenarios show how the various performance measurement techniques may be applied to the measurement network models. The flexibility of the models support more applied measurements than those described previously.

In the following scenarios, the measurement information exchanged among providers every rollup period includes the following:

1) minimum delay;
2) mean delay;
3) high delay percentiles;
4) loss ratio;
5) unavailability period information;
6) miscellaneous information.

All measurement scenarios described below are applicable to active, passive and spatial measurement techniques, unless otherwise noted. When measurement results have been obtained, the results should be conveyed from the collection points to management systems with oversight responsibility. Appendix II gives a description of a generic management process for measurement systems. However, the details of the management process are beyond the scope of this Recommendation.

7.2.1 Operating measurements scenario

The site-site operating measurement scenario is shown in Figure 4, where the endpoints are CEs.

Figure 4 represents two connected service providers, A and B, each having regional and access networks on which end customers have managed CEs. The following operating measurements are needed to estimate the site-to-site performance being delivered for customer site A.

1) SP A initiates measurements between:
   a) DP A and site A CE; and
   b) DP A and DP B.
2) SP B initiates measurements between:
   a) DP B and site B CE.
3) SP A retrieves results of measurements between:
   a) DP B and site B CE from SP B.
4) SP A compares the aggregated metric of the 3 segments to the guarantee and provides a report to site A customer.

The supporting measurements for the above service are detailed in Figure 6.

The edge-edge operating measurement scenario is shown in Figure 3, where the endpoints are PEs.

Similar to Figure 4, the following operating measurements are required to estimate the edge-to-edge performance being delivered for customer site A. Note that the only difference to Figure 4 is the use of PEs versus the use of CEs.

1) SP A initiates measurements between:
   a) DP A and site A PE; and
   b) DP A and DP B.
2) SP B initiates measurements between:
   a) DP B and site B PE.
3) SP A retrieves results of measurements between:
   a) DP B and site B PE from SP B.
4) SP A compares the aggregated metric of the 3 segments to the guarantee and provides a report to site A customer.

The supporting measurements are similar to those shown in Figure 6.

The TE-TE operating measurement scenario is shown in Figure 5, where the endpoints are TEs.

Similar to Figure 4, the following operating measurements are required to estimate the TE-to-TE performance being delivered for customer A. Note that the only difference to Figure 4 is the addition of CE-TE measurements and the retrieval of those measurements from SP B.

1) SP A initiates measurements between:
   a) DP A and site A CE;
   b) Site A CE and Site A TE;

    c) DP A and DP B.

2) SP B initiates measurements between:

    a) DP B and site B CE;

    b) Site B CE and site B TE.

3) SP A retrieves results of measurements from SP B for measurements between:

    a) DP B and site B CE;

    b) Site B CE and site B TE.

4) SP A compares the aggregated metric of the 5 segments to the guarantee and provides a report to site A customer.

This scenario assumes a single TE; measurements with multiple TEs may be supported.

7.2.2 Supporting measurements scenario

To provide measurements for purposes in the support category, SPs may choose to perform measurement across their network to key measurement points in other cities/regions of their networks. SPs may choose the mechanism used for internal support measurements, which may be the same as used for operating measurements. It is recommended, however, that each SP implement sufficient support tools to enable resolution of performance issues within their networks.

In many cases, the sending and receiving customer sites will be connected to the same regional service provider. To support these cases, the regional service provider is fully responsible for the transit segment of the network and should perform the appropriate measurement functions.

The transit segment of the network will often comprise of two regional SPs and one backbone SP and their interconnections. Each of these service providers should enable resolution of performance issues that may occur. This resolution process will include monitoring of specific sections of the transit segment. The collection of these statistics is part of the support process.

An SP should measure performance to each neighbouring POP of the other directly connected regional SPs and backbone SPs. This enables issue resolution of the interconnect performance and dimensioning as opposed to the network performance of the neighbouring SPs. In some cases, the interconnects may be through peering points with more complex performance characteristics and in other cases, high speed SONET/SDH interconnect may be used. The interconnect egress performance and dimensioning is the responsibility of the regional SP that the customer connects to. In the case of a regional SP interconnect to a backbone SP, the performance and dimensioning of both directions through the interconnect is the responsibility of the regional SP. The backbone SP may supply services that simplify this process and ensure performance targets are met.

In Figure 6 below, SP A and B each measure their own contribution to the transit segment, and may allow other interconnected SPs to retrieve those measurements. In this case, SP A and SP B each includes in their measurements one direction through the interconnect. SP A may retrieve the measurements of SP B's contribution to the transit segment.

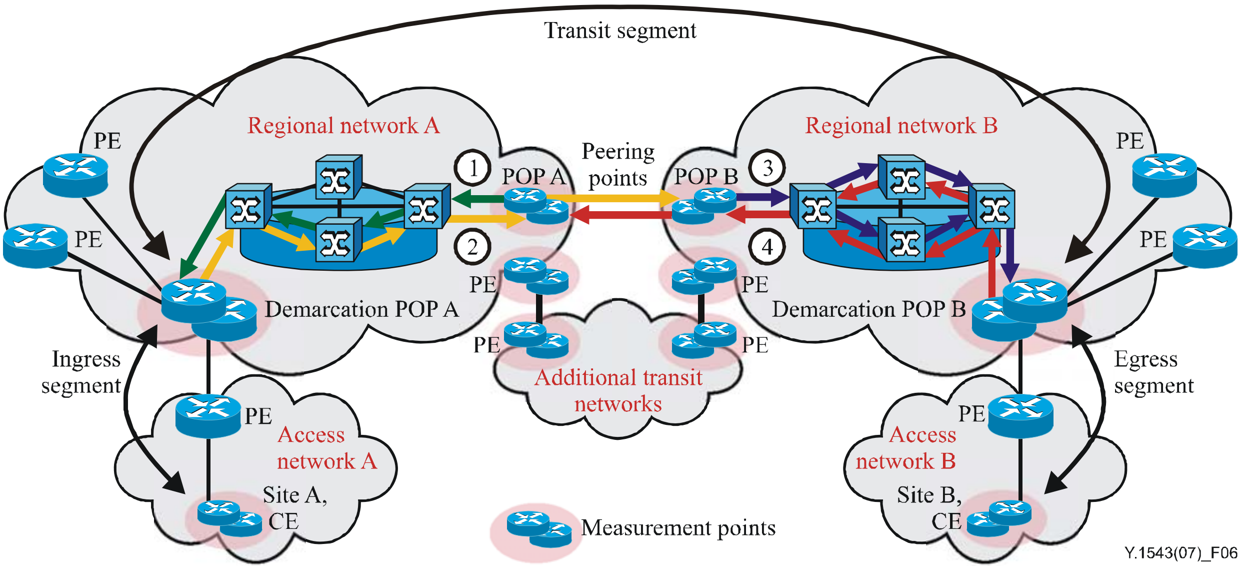


Figure 6 – Site-site supporting measurements

In this scenario, assume that each SP is responsible for performance assurance of their egress traffic over peering point links.

In order to obtain supporting measurements for the customer service indicated in Figures 4 and 3, the following activities would be performed as indicated above in Figure 6:

1) SP A initiates measurements between DP A and its peering point POP A, for the direction from peering point POP A to DP A.

2) SP A initiates measurements between DP A and network B peering point POP, for the direction DP A to peering point POP B.

3) SP B initiates measurements between DP B and its peering point POP B, for the direction from peering point POP B to DP B.

4) SP B initiates measurements between DP B and network A peering point POP, for the direction from DP B to peering point POP A.

In addition to this data being used for an SP to confirm its own transit performance, these measurements may be concatenated:

- If aggregated with DP-CE measurements it is an estimator of a SP's total CE-peering point performance.
- This data may be exchanged with partner SPs to provide assurance, and if aggregated with other supporting measurements, may be used as a sanity check for operating measurements.

Note that in Figure 7, four additional supporting measurements are required, 2 for each SP's part of the transit segment. Further addition of IDQ services across this network would not require additional transit segment measurements but would reuse the results of these measurements. Extension of the model in Figure 7 to include more SPs would require additional supporting measurement of that SP's transit segment.

The description for supporting measurements above is for the case where each SP is responsible for its egress traffic over a single peering link. Similar scenarios may be used in the case of:

1) dual links (in parallel) where each provider pays for one of the links, and both links are actively used;

2) third party internet exchange points.

7.2.3 Test measurements scenario

Information useful for troubleshooting or prospective customers may require additional measurements.

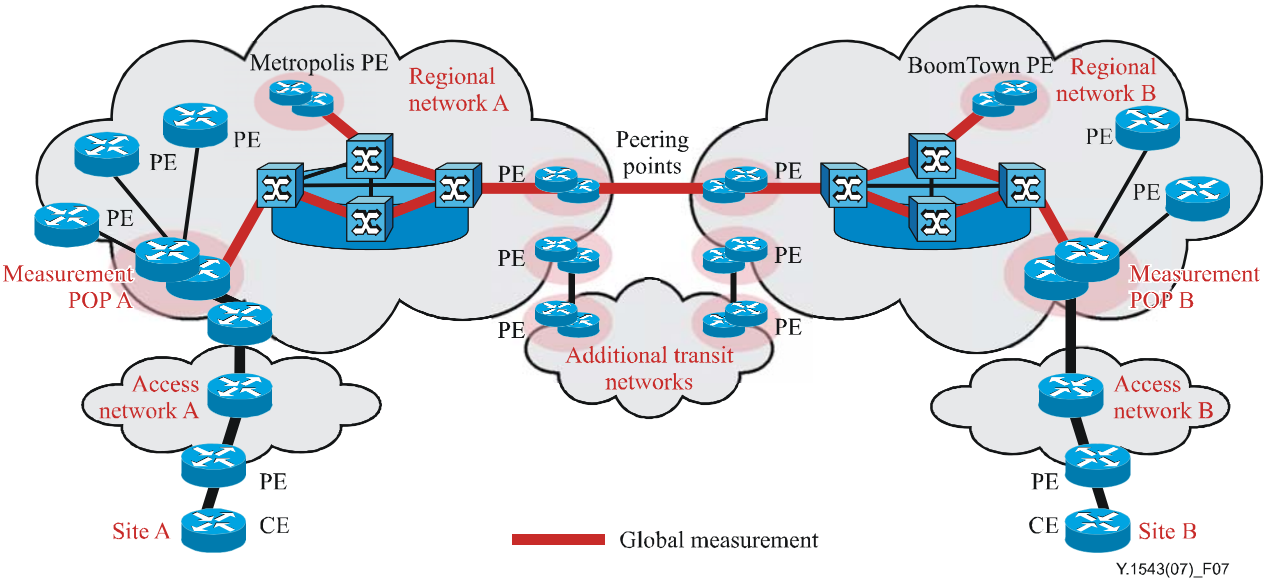


Figure 7 – Global-global measurements

In Figure 7, SP A or SP B initiates measurements between major global POPs named Metropolis POP and BoomTown POP, and publishes them in a report. This report indicates whether the transit performance targets are being met for a significant set of destinations and approximates the expected performance for nearby POPs. This is useful for SP A and SP B as a basis for offering prospective services to customers who connect through or close to the Metropolis and BoomTown POPs.

SP A may wish to initiate measurement between DP A and the Metropolis measurement POP. This may be useful for SP A as a basis for offering prospective services to customers who connect both source and destination CEs to SP A's network.

Measurements from each demarcation POP to a significant set of high profile, global measurement POPs of multiple SPs may occur for similar purposes. This set of measurements characterizes the transit segment of the network for a representative set of customer traffic flows. The selected global POPs should cover all major cities and continents and include many other service providers. It is expected that a minimum of 50 global POPs would be monitored from each demarcation POP.

In some cases, a customer may not be satisfied that any of the chosen set of global measurement POPs is sufficient to characterize a specific transit segment. At a customer's request, an SP may initiate measurements between the customer's DP and a set of selected POPs. This would normally be viewed as a custom service. Along with custom end-points, additional statistics and reports could be provided.

7.2.4 Example RTP/RTCP-based passive measurement scenario

Most real-time multimedia applications on IP-based networks use RTP. RTP/RTCP-based passive measurement is effective in that it can assist to collect network performance. A per-segment based measurement scenario is shown below in Figure 8, where the measurement points are TEs and border gateways (BGs) which handle RTP/RTCP and RTCP extension packets.

Delay and delay-variation are important to real-time applications such as VoIP and video-streaming. Real time protocol, [IETF RFC 3550]) is a transport layer protocol for real-time applications. RTP is designed to be independent of transport or network layer protocols. An RTP packet has time-stamp and sequence-number fields in its header. A passive collection system which resides in either TEs or BGs can evaluate packet loss and delay variation.

RTCP is an optional control protocol for RTP. Furthermore, RTCP extensions, such as the RTCP-XR (RTP control protocol extended reports, [b-IETF RFC 3611]) are also optional control protocols for RTP. Participating TEs exchange RTCP and RTCP-XR packets. In a RTCP and RTCP-XR packet, performance metrics of its application services are reported. TEs also are able to evaluate rough round-trip delay with these packets.

Figure 8 represents two connected service providers, A and B, each having regional and access networks to which end customers have managed TEs.

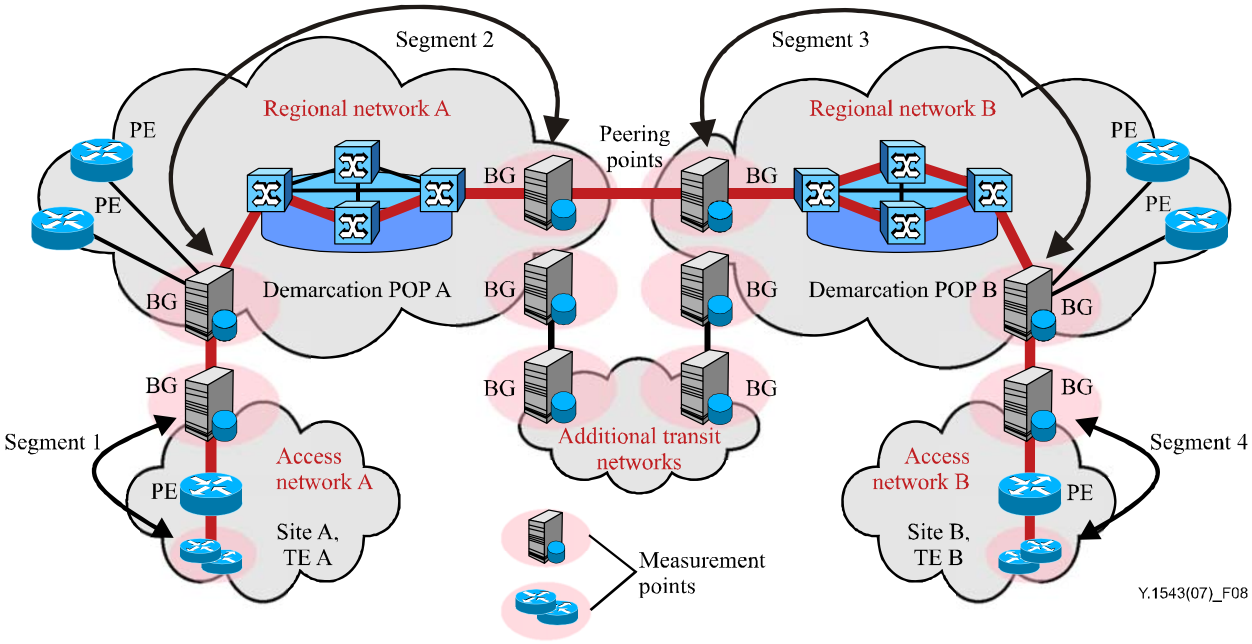


Figure 8 – RTP/RTCP based passive measurement

The following operating measurements are required to estimate the per-segment based performance between TE A and TE B using RTP/RTCP and RTCP extension packets:

1) TE A initiates RTP sessions between TE A of site A and TE B of site B.

2) When the sessions are established, TE A and TE B communicate RTP/RTCP packets as well as RTCP-XR packets.

3) BGs which are located over the path of the RTP/RTCP flows measure the performance based on the information provided in RTP/RTCP/RTCP-XR packets.

4) TEs and BGs can independently collect the performance metrics of their own segment.

5) The reports of metrics collected at TEs and BGs are sent to the management system.

The TE-to-TE, site-to-site, edge-to-edge scenarios do not apply in the case of RTP/RTCP-based passive measurement procedures since it is an application level performance measurement. It only has significance in terms of the TE-to-TE case. However, the performance metrics of other scenarios (site-to-site and edge-to-edge) can still be reported as described by the above procedure.

NOTE 1 – For conversational applications, metric reports of both directions between TE A and TE B are collected at BGs at the same time.

NOTE 2 – For one-way streaming applications, metrics reports of the selected direction between TE A and TE B are collected at BGs.

NOTE 3 – Management systems can choose TEs and BGs which send metrics reports depending on the type of application.

7.2.5 Example spatial measurement scenario

Active and passive measurement procedures can be used independently to meet specific purposes. However, there may be a situation where passive measurements can take advantage of active probes enabling both methods to be used cooperatively. For example, a TE-to-TE operating active measurement procedure requires separate measurement of several segments, and each segment requires a pair of active measurement points. The measurement results from each segment are aggregated to make end-to-end metrics. Another possible solution is to have a single TE-to-TE active measurement and a number of passive measurement devices deployed at some MPs such as demarcation POPs, PEs and CEs. Each passive collection system can recognize the target active probe packets and measure network level performance metrics from the initiating TE to the passive measurement points. For this method to be possible, we use Performance Measurement Specification (IPPMS) which is defined in [ITU-T O.211] to uniquely identify the active probe packet across multiple administration domains. It defines controller ID and flow ID to make an unique identification across multiple administration domains. The passive measurement devices which can recognize IPPMS capture the desired packets for various measurement operations.

The main advantage of this method is to reduce a large number of active probes in the middle of managed networks since one passive measurement device can handle a large number of active measurement sessions. The procedure below gives one possible measurement scenario (refer to Figure 5 and Appendix II).

The following operating measurements are required to estimate the performance between TE A and TE B by hybrid methods of active and passive measurement.

1) SP A initiates active inter-domain measurement between TE A and TE B:
   a) TE A sends active probe packets (conformant to [ITU-T O.211]) to TE B.
2) SP A initiates passive measurements by observing target probe packets:
   a) Site A CE, PE, DP A, and peering point A collect target packets.
   b) Site A CE, PE, DP A, and peering point A report to its management system.
   c) Management system A generates performance metrics of each segment.
3) SP B initiates measurements by observing target probe packets:
   a) Site B peering point B, DB B, PE, and CE collect target packets.
   b) Site B peering point B, DB B, PE, and CE report to its management system.
   c) Management system B generates performance metrics of each segment.
4) SP A and B can exchange and compare metrics to generate customer reports.

Besides TE-to-TE measurement, site-to-site and edge-to-edge measurement can be similarly performed. Thus, their procedures are not repeated.

8 Measurement procedures

8.1 Active measurement procedures

8.1.1 Mean one-way delay

The delay attributes of a network QoS class over a network segment are characterized by minimum delay, mean delay and a specific set of upper percentile delays. The percentile approach is used in preference to a standard deviation or variance model due to the frequent occurrence of bi-modal or multi-modal delay distributions.

In real networks, there are occasional events such as rerouting and momentary link outages that cause significant additional delays over and above the normal propagation and queuing delays. Packets that are delayed excessively are of little or no value to the application being supported and could be treated as lost packets; however, the incremental value of doing so is not considered to be worth the additional complexity, therefore delay outliers will be included in the delay statistics.

The segment one-way mean delay is calculated as follows:

1) Collect measurements from N probes generated every probe transmission period (PTP) for each rollup period (RP).

2) Discard all measurements from periods of unavailability.

3) Mean delay = sum(1..M) measurements/M (where M is the number of successful packet transfers, possibly less than N).

Multi-segment mean delay is calculated by aggregating the mean delays of each segment mean delay through a simple summation.

Measurement samples from unavailability periods are not included in statistics.

8.1.2 One-way delay variation

Segment (one-way) delay variation (DV) is derived from the minimum delay and percentile. It is derived on a rollup period basis. For each segment,

$$\text{DV} = \text{One-way\_Delay\_Percentile} - \text{Minimum}$$

For specific percentiles,

$$\text{DV99.9} = \text{99.9Percentile} - \text{Minimum}$$

$$\text{DV99} = \text{99Percentile} - \text{Minimum}$$

$$\text{DV90} = \text{90Percentile} - \text{Minimum}$$

Multiple segment delay variations are used per network QoS class as follows:

Table 2 – Segment delay variations used per network QoS class

DV	Most stringent QoS class	Mid-level stringency QoS class	Least stringent QoS class
DV99.9	x		
DV99	x	x	
DV90	x	x	x

Segment one-way delay percentiles are calculated as follows:

1) Collect measurements from N probes generated every probe transmission period (PTP) for each rollup period (RP).
2) Discard all measurements from periods of loss or unavailability, leaving M samples.
3) Stack rank the measurement set.
4) Discard the top D measurements (D = round((100 – percentile) × M)).
5) Percentile = delay value of top remaining sample.

Multi-segment delay variation is calculated by aggregating the delay variations of each segment through a provisional method defined in [ITU-T Y.1541]. It is also derived on a rollup period basis.

Since minimum delay and percentiles from unavailability periods are not included in statistics, derived DVs are also not included from unavailability periods.

8.1.3 Packet loss ratio

Segment (one-way) packet loss (PL) is measured over the same period as delay. It is derived on a rollup period basis. Segment packet loss is the number of probes whose measured one-way delay was $\geq T_{max}$ (N_T_{max}) and those that never made it to their destination, or missing (MSNG). Note that T_{max} is the waiting time defined in [ITU-T Y.1540].

$$PL = (N\_T_{max} + MSNG)$$

Packet loss ratio (PLR) is packet loss divided by the number of transmitted packets (N)

$$PLR = PL/N = (N\_T_{max} + MSNG)/N$$

Measurements from unavailability periods are not included in packet loss statistics. Both the number of lost packets and the number of transmitted packets are reduced accordingly. This process avoids the packet loss ratios being unduly impacted by network unavailability.

To combine these to produce a multi-segment packet loss ratio, called the aggregate loss ratio (ALR), the following method is used.

$$ALR = 1 - (1 - \text{PLR for segment 1}) \times (1 - \text{PLR for segment 2}) \times (1 - \text{PLR for segment 3})$$

ALR is derived for each rollup period.

8.1.4 Path unavailability

Unavailability is determined on a per QoS class, per direction basis, from one-way packet loss measurements. The Y.1540 service availability function is the basis for measurement of unavailability.

The window for availability evaluation (T_{av}) should be aligned with the rollup period (5 minutes). The loss threshold for state determination is $c_1 = 0.75$ (75% packet loss ratio).

In order to calculate one-way unavailability, the absence of a one-way delay measurement must be understood to be due to an outbound loss rather than an inbound loss.

Delay, delay variation and loss measurements and their derived metrics are ignored for a segment for the duration of its unavailability.

For each segment, unavailability is calculated by summing all the rollup periods determined to be in the unavailable state.

Multi-segment path unavailability is calculated by calculating the total of unavailable time by adding the non-overlapping unavailable rollup periods from each segment in the path.

8.2 Passive measurement procedures

These procedures follow the general inter-domain measurement procedure, and specifics regarding passive measurement functions at the measurement points are the only difference.

Two passive measurement collection systems extract flow summary data (FSD) from the packets of target flow and attach time-stamps. The collection system may be dedicated hardware tapping the optical signal from the transmission link, or may be a software or hardware module installed in a network element. FSDs from the two collection systems are sent to each management system. Single probe results are compiled by the initiating management system. Unlike active measurement procedures, the path of target flow may change during the measurement period. In such a case, one or both of the collection systems may not be able to extract the necessary FSDs. This should be perceived by the management systems and appropriate measurement actions such as relocating MPs and restarting the measurement activity.

9 Areas for further study and future work

Recommendations:

a) Develop accuracy requirements for measured values and their implications.

10 Security considerations

10.1 Impact of security on measurement of performance

The strength of security measures used in a solution can burden systems, and/or cause extra security-related traffic. Since a heavily burdened router or firewall, or waiting for security-related traffic to return, may delay measurements, some risks versus benefits need to be considered.

To meet the high level of security requirements listed above by implementing authentication and data integrity into the probes would require additional overhead on the measurement devices to do the authentication and data integrity. Depending on the number of probes, this could impact the measurement devices with the overhead caused by this operation.

To handle the performance of doing authentication, this could be done on the measurement device itself or off-loaded to another system. The recommendation is to do the authentication on the measurement device itself since it would likely allow for faster response than off-loading to another device requiring security-related network traffic.

10.2 Impact of performance measurement on security

User traffic may be collected as a result of passive measurements. Since user payloads may be temporarily stored for later analysis, suitable precautions must be taken to keep this information safe and confidential.

Appendix I

Summary of performance objectives and measurements

(This appendix does not form an integral part of this Recommendation)

Table I.1 – Summary of performance objectives and measurements

Network parameter acronym	**Parameter description**	**Performance objective**	**Units**	**Y.1541 network QoS classes**								**Relative to Y.1541**	**Covered in clauses**
				Class 0	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**	**Class 7**		
IPTD	Mean IP packet transfer delay	Upper bound over RP	ms	100	400	100	400	1.000	U	100	400	Per Y.1541. Classes 1 and 3 are for less constrained distance than 0 and 2 respectively	5.1, 8.1.1
DV90	IP packet delay variation 90th percentile – minimum IPTD	Upper bound on delay variation over RP	ms	future	future	future	future	future	future	future	future	Not covered in Y.1541	5.2, 8.1.2
DV99	IP packet delay variation 99th percentile – minimum IPTD	Upper bound on delay variation over RP	ms	future	future	future	future	future	future	future	future	Not covered in Y.1541	5.2, 8.1.2
IPDV, DV99.9	IP packet delay variation 99.9th percentile – minimum IPTD	Upper bound on delay variation over RP	ms	50	50	U	U	U	U	50	50	Per Y.1541	5.2, 8.1.2
IPLR, ALR	IP packet loss ratio, aggregate loss ratio	Upper bound on the packet loss probability over RP	%	0.1	0.1	0.1	0.1	0.1	U	0.001	0.001	Per Y.1541	5.3, 8.1.3

Table I.1 – Summary of performance objectives and measurements

Network parameter acronym	**Parameter description**	**Performance objective**	**Units**	**Y.1541 network QoS classes**								**Relative to Y.1541**	**Covered in clauses**
				Class 0	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**	**Class 6**	**Class 7**		
IPUA	Total period of excessive short term loss during which the network is considered unavailable	Upper bound on the percentage over month	%	0	0	0	0	0	1	0	0	Not covered in Y.1541	5.4, 8.1.4
PW	Policing window	Corollary	Minutes	5	5	5	5	5	5	5	5	Not covered in Y.1541	6.5
PPS	Probe payload size(s)	Corollary	Octets	20	20	256	256	256	256	20	20	Y.1541 suggests 160 or 1500 octets. Per Y.1541 proposal.	6.1.2
RP	Rollup period	Corollary	Minutes	5	5	5	5	5	5	5	5	Y.1541 suggests 1 minute. Per Y.1541 proposal.	6.1.2
PTP	Probe transmission period (continuous)	Corollary	ms	100	100	100	100	100	100	100	100	Y.1541 suggests 10 to 20 ms for telephony. Y.1541 proposal suggests 20 ms for classes 0, 1 or 50 ms for classes 2, 3, 4 for 1 minute sampled out of 5 minutes	6.1.2

NOTE 1 – Y.1541 in this table refers to [ITU-T Y.1541].

NOTE 2 – Classes 6 and 7 are provisional classes.

NOTE 3 – U means unspecified (unbounded).

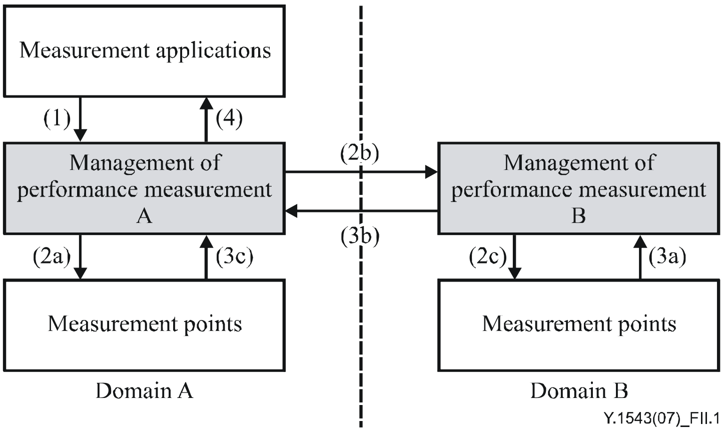
NOTE 4 – Regarding the Y.1541 network QoS classes, note that certain pairs of classes collapse into node mechanisms/per hop behaviours/queues. These are classes 0 and 1, 2 and 3, and 6 and 7. Therefore, to probe all classes would require four measurement flows. Measurement results for these pairs would be used in comparison to objectives for each class.

Appendix II

Generic inter-domain management process for measurement systems

(This appendix does not form an integral part of this Recommendation)

Figure II.1 below depicts the general procedures for management of inter-domain performance measurement systems. Measurement point (MP) is a functional entity located in the transport, transport control, or service control networks. In case of active measurement, it is responsible for initiating and receiving probe packets. In case of passive measurement, it is responsible for capturing target packets. Management of performance measurement (MPM) functions include the interaction with measurement applications and the MPs, configuration of MPs, and exchanging the required configuration and measured information. The details of this process are a topic of active study in ITU-T SG 13. The following procedures are described based on such capabilities.



(1) The measurement application of SP A initiates a measurement task by sending measurement request to MPM.

(2a) Upon receipt of measurement request, MPM A locates the involved MPs. For the MPs located in domain A, MPM A sends the measurement parameters to MPs.

(2b) Upon receipt of measurement request, MPM A locates the involved MPs. For the MPs located in domain B, MPM A sends the measurement request to MPM B.

(2c) Upon receipt of measurement request, MPM B locates the involved MPs. For the MPs located in domain B, MPM B sends the measurement parameters to MPs.

(3a) MPM B collects the measured data from MPs located in domain B.

(3b) MPM B sends the measurement information to MPM A.

(3c) MPM A collects the measured data from MPs located in domain A.

(4) Based on the received measurement information from domain A and domain B, MPM A sends the response to the measurement applications.

Figure II.1 – Generic inter-domain management process

Bibliography

[b-ITU-T G.810] Recommendation ITU-T G.810 (1996), *Definitions and terminology for synchronization networks*.

[b-ITU-T I.356] Recommendation ITU-T I.356 (2000), *B-ISDN ATM layer cell transfer performance*.

[b-IETF RFC 1889] IETF RFC 1889 (1996), *RTP: A Transport Protocol for Real-Time Applications*. <http://www.ietf.org/rfc/rfc1889.txt?number=1889>

[b-IETF RFC 2678] IETF RFC 2678 (1999), *IPPM Metrics for Measuring Connectivity* <http://www.ietf.org/rfc/rfc2678.txt?number=2678>.

[b-IETF RFC 3611] IETF RFC 3611 (2003), *RTP Control Protocol Extended Reports (RTCP XR)* <http://www.ietf.org/rfc/rfc3611.txt?number=3611>.

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Telephone transmission quality, telephone installations, local line networks
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects and next-generation networks**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2008

Network Working Group
Request for Comments: 2678
Obsoletes: 2498
Category: Standards Track

J. Mahdavi
Pittsburgh Supercomputing Center
V. Paxson
Lawrence Berkeley National Laboratory

September 1999

IPPM Metrics for Measuring Connectivity

Status of this Memo

This document specifies an Internet standards track protocol for the Internet community, and requests discussion and suggestions for improvements. Please refer to the current edition of the "Internet Official Protocol Standards" (STD 1) for the standardization state and status of this protocol. Distribution of this memo is unlimited.

Copyright Notice

1. Introduction

Connectivity is the basic stuff from which the Internet is made. Therefore, metrics determining whether pairs of hosts (IP addresses) can reach each other must form the base of a measurement suite. We define several such metrics, some of which serve mainly as building blocks for the others.

This memo defines a series of metrics for connectivity between a pair of Internet hosts. It builds on notions introduced and discussed in RFC 2330, the IPPM framework document. The reader is assumed to be familiar with that document.

The structure of the memo is as follows:

+ An analytic metric, called Type-P-Instantaneous-Unidirectional-

Connectivity, will be introduced to define one-way connectivity at one moment in time.

+ Using this metric, another analytic metric, called Type-P-Instantaneous-Bidirectional-Connectivity, will be introduced to define two-way connectivity at one moment in time.
+ Using these metrics, corresponding one- and two-way analytic metrics are defined for connectivity over an interval of time.

+ Using these metrics, an analytic metric, called Type-P1-P2-Interval-Temporal-Connectivity, will be introduced to define a useful notion of two-way connectivity between two hosts over an interval of time.
+ Methodologies are then presented and discussed for estimating Type-P1-P2-Interval-Temporal-Connectivity in a variety of settings.

Careful definition of Type-P1-P2-Interval-Temporal-Connectivity and the discussion of the metric and the methodologies for estimating it are the two chief contributions of the memo.

2. Instantaneous One-way Connectivity

2.1. Metric Name:

Type-P-Instantaneous-Unidirectional-Connectivity

2.2. Metric Parameters:

+ Src, the IP address of a host
+ Dst, the IP address of a host
+ T, a time

2.3. Metric Units:

Boolean.

2.4. Definition:

Src has *Type-P-Instantaneous-Unidirectional-Connectivity* to Dst at time T if a type-P packet transmitted from Src to Dst at time T will arrive at Dst.

2.5. Discussion:

For most applications (e.g., any TCP connection) bidirectional connectivity is considerably more germane than unidirectional connectivity, although unidirectional connectivity can be of interest for some security applications (e.g., testing whether a firewall correctly filters out a "ping of death"). Most applications also require connectivity over an interval, while this metric is instantaneous, though, again, for some security applications instantaneous connectivity remains of interest. Finally, one might not have instantaneous connectivity due to a transient event such as a full queue at a router, even if at nearby instants in time one does have connectivity. These points are addressed below, with this metric serving as a building block.

Note also that we have not explicitly defined *when* the packet arrives at Dst. The TTL field in IP packets is meant to limit IP packet lifetimes to 255 seconds (RFC 791). In practice the TTL field can be strictly a hop count (RFC 1812), with most Internet hops being much shorter than one second. This means that most packets will have nowhere near the 255 second lifetime. In principle, however, it is also possible that packets might survive longer than 255 seconds. Consideration of packet lifetimes must be taken into account in attempts to measure the value of this metric.

Finally, one might assume that unidirectional connectivity is difficult to measure in the absence of connectivity in the reverse direction. Consider, however, the possibility that a process on Dst's host notes when it receives packets from Src and reports this fact either using an external channel, or later in time when Dst does have connectivity to Src. Such a methodology could reliably measure the unidirectional connectivity defined in this metric.

3. Instantaneous Two-way Connectivity

3.1. Metric Name:

Type-P-Instantaneous-Bidirectional-Connectivity

3.2. Metric Parameters:

+ A1, the IP address of a host
+ A2, the IP address of a host
+ T, a time

3.3. Metric Units:

Boolean.

3.4. Definition:

Addresses A1 and A2 have *Type-P-Instantaneous-Bidirectional-Connectivity* at time T if address A1 has Type-P-Instantaneous-Unidirectional-Connectivity to address A2 and address A2 has Type-P-Instantaneous-Unidirectional-Connectivity to address A1.

3.5. Discussion:

An alternative definition would be that A1 and A2 are fully connected if at time T address A1 has instantaneous connectivity to address A2, and at time T+dT address A2 has instantaneous connectivity to A1, where T+dT is when the packet sent from A1 arrives at A2. This definition is more useful for measurement, because the measurement can use a reply from A2 to A1 in order to assess full connectivity. It is a more complex definition, however, because it breaks the symmetry between A1 and A2, and requires a notion of quantifying how long a particular packet from A1 takes to reach A2. We postpone discussion of this distinction until the development of interval-connectivity metrics below.

4. One-way Connectivity

4.1. Metric Name:

Type-P-Interval-Unidirectional-Connectivity

4.2. Metric Parameters:

+ Src, the IP address of a host
+ Dst, the IP address of a host
+ T, a time
+ dT, a duration

{Comment: Thus, the closed interval [T, T+dT] denotes a time interval.}

4.3. Metric Units:

Boolean.

4.4. Definition:

Address Src has *Type-P-Interval-Unidirectional-Connectivity* to address Dst during the interval [T, T+dT] if for some T' within [T, T+dT] it has Type-P-instantaneous-connectivity to Dst.

5. Two-way Connectivity

5.1. Metric Name:

Type-P-Interval-Bidirectional-Connectivity

5.2. Metric Parameters:

+ A1, the IP address of a host

+ A2, the IP address of a host
+ T, a time
+ dT, a duration

{Comment: Thus, the closed interval [T, T+dT] denotes a time interval.}

5.3. Metric Units:

Boolean.

5.4. Definition:

Addresses A1 and A2 have *Type-P-Interval-Bidirectional-Connectivity* between them during the interval [T, T+dT] if address A1 has Type-P-Interval-Unidirectional-Connectivity to address A2 during the interval and address A2 has Type-P-Interval-Unidirectional-Connectivity to address A1 during the interval.

5.5. Discussion:

This metric is not quite what's needed for defining "generally useful" connectivity - that requires the notion that a packet sent from A1 to A2 can elicit a response from A2 that will reach A1. With this definition, it could be that A1 and A2 have full-connectivity but only, for example, at time T1 early enough in the interval [T, T+dT] that A1 and A2 cannot reply to packets sent by the other. This deficiency motivates the next metric.

6. Two-way Temporal Connectivity

6.1. Metric Name:

Type-P1-P2-Interval-Temporal-Connectivity

6.2. Metric Parameters:

- Src, the IP address of a host
- Dst, the IP address of a host
- T, a time
- dT, a duration

{Comment: Thus, the closed interval [T, T+dT] denotes a time interval.}

6.3. Metric Units:

Boolean.

6.4. Definition:

Address Src has *Type-P1-P2-Interval-Temporal-Connectivity* to address Dst during the interval [T, T+dT] if there exist times T1 and T2, and time intervals dT1 and dT2, such that:

- T1, T1+dT1, T2, T2+dT2 are all in [T, T+dT].
- T1+dT1 <= T2.
- At time T1, Src has Type-P1 instantanous connectivity to Dst.
- At time T2, Dst has Type-P2 instantanous connectivity to Src.
- dT1 is the time taken for a Type-P1 packet sent by Src at time T1 to arrive at Dst.
- dT2 is the time taken for a Type-P2 packet sent by Dst at time T2 to arrive at Src.

6.5. Discussion:

This metric defines "generally useful" connectivity -- Src can send a packet to Dst that elicits a response. Because many applications utilize different types of packets for forward and reverse traffic, it is possible (and likely) that the desired responses to a Type-P1 packet will be of a different type Type-P2. Therefore, in this metric we allow for different types of packets in the forward and reverse directions.

6.6. Methodologies:

Here we sketch a class of methodologies for estimating Type-P1-P2-Interval-Temporal-Connectivity. It is a class rather than a single methodology because the particulars will depend on the types P1 and P2.

6.6.1. Inputs:

+ Types P1 and P2, addresses A1 and A2, interval [T, T+dT].
+ N, the number of packets to send as probes for determining connectivity.
+ W, the "waiting time", which bounds for how long it is useful to wait for a reply to a packet.

Required: W <= 255, dT > W.

6.6.2. Recommended values:

dT = 60 seconds.
W = 10 seconds.
N = 20 packets.

6.6.3. Algorithm:

+ Compute N *sending-times* that are randomly, uniformly distributed over [T, T+dT-W].
+ At each sending time, transmit from A1 a well-formed packet of type P1 to A2.
+ Inspect incoming network traffic to A1 to determine if a successful reply is received. The particulars of doing so are dependent on types P1 & P2, discussed below. If any successful reply is received, the value of the measurement is "true". At this point, the measurement can terminate.
+ If no successful replies are received by time T+dT, the value of the measurement is "false".

6.6.4. Discussion:

The algorithm is inexact because it does not (and cannot) probe temporal connectivity at every instant in time between [T, T+dT]. The value of N trades off measurement precision against network measurement load. The state-of-the-art in Internet research does not yet offer solid guidance for picking N. The values given above are just guidelines.

6.6.5. Specific methodology for TCP:

A TCP-port-N1-port-N2 methodology sends TCP SYN packets with source port N1 and dest port N2 at address A2. Network traffic incoming to A1 is interpreted as follows:

+ A SYN-ack packet from A2 to A1 with the proper acknowledgement fields and ports indicates temporal connectivity. The measurement terminates immediately with a value of "true". {Comment: if, as a side effect of the methodology, a full TCP connection has been established between A1 and A2 -- that is, if A1's TCP stack acknowledges A2's SYN-ack packet, completing the three-way handshake -- then the connection now established between A1 and A2 is best torn down using the usual FIN handshake, and not using a RST packet, because RST packets are not reliably delivered. If the three-way handshake is not completed, however, which will occur if the measurement tool on A1 synthesizes its own initial SYN packet rather than going through A1's TCP stack, then A1's TCP stack will automatically terminate the connection in a reliable fashion as A2 continues transmitting the SYN-ack in an attempt to establish the connection. Finally, we note that using A1's TCP stack to conduct the measurement complicates the methodology in that the stack may retransmit the initial SYN packet, altering the number of probe packets sent.}

+ A RST packet from A2 to A1 with the proper ports indicates temporal connectivity between the addresses (and a *lack* of service connectivity for TCP-port-N1-port-N2 - something that probably should be addressed with another metric).
+ An ICMP port-unreachable from A2 to A1 indicates temporal connectivity between the addresses (and again a *lack* of service connectivity for TCP-port-N1-port-N2). {Comment: TCP implementations generally do not need to send ICMP port-unreachable messages because a separate mechanism is available (sending a RST). However, RFC 1122 states that a TCP receiving an ICMP port-unreachable MUST treat it the same as the equivalent transport-level mechanism (for TCP, a RST).}
+ An ICMP host-unreachable or network-unreachable to A1 (not necessarily from A2) with an enclosed IP header matching that sent from A1 to A2 *suggests* a lack of temporal connectivity. If by time T+dT no evidence of temporal connectivity has been gathered, then the receipt of the ICMP can be used as additional information to the measurement value of "false".

{Comment: Similar methodologies are needed for ICMP Echo, UDP, etc.}

7. Acknowledgments

The comments of Guy Almes, Martin Horneffer, Jeff Sedayao, and Sean Shapira are appreciated.

8. Security Considerations

As noted in RFC 2330, active measurement techniques, such as those defined in this document, can be abused for denial-of-service attacks disguised as legitimate measurement activity. Furthermore, testing for connectivity can be used to probe firewalls and other security mechnisms for weak spots.

9. References

[RFC1812] Baker, F., "Requirements for IP Version 4 Routers", RFC 1812, June 1995.

[RFC1122] Braden, R., Editor, "Requirements for Internet Hosts -- Communication Layers", STD, 3, RFC 1122, October 1989.

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

[RFC791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.

10. Authors' Addresses

Jamshid Mahdavi
Pittsburgh Supercomputing Center
4400 5th Avenue
Pittsburgh, PA 15213
USA

EMail: mahdavi@psc.edu

Vern Paxson
MS 50A-3111
Lawrence Berkeley National Laboratory
University of California
Berkeley, CA 94720
USA

Phone: +1 510/486-7504

EMail: vern@ee.lbl.gov

11. Full Copyright Statement

This document and the information contained herein is provided on an "AS IS" basis and THE INTERNET SOCIETY AND THE INTERNET ENGINEERING TASK FORCE DISCLAIMS ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.