Network Working Group
INTERNET-DRAFT
Expires in: December 2001

Jerry Perser
Spirent
David Newman
Network Test
Sumit Khurana
Telcordia
Shobha Erramilli
Telcordia
Scott Poretsky
Avici Systems
June 2001

Terminology for Benchmarking Network-layer
Traffic Control Mechanisms

<draft-ietf-bmwg-dsmterm-01.txt>

Status of this Memo

Table of Contents

1. Introduction

Driven by Internet economics, service providers and enterprises alike have shown strong interest in adding traffic-control capabilities to network devices. These capabilities would enable network operators to define and deliver minimum or maximum levels of bandwidth, delay, and jitter for multiple classes of traffic. Perhaps more importantly, network operators would be able to set pricing according to the level of service delivered.

Networking device manufacturers have responded with a wide variety of approaches for controlling network traffic. While there are numerous ôpolicy managementö and ôquality of serviceö frameworks, many of them rely on one of two network-layer mechanisms for the actual control of forwarding rate, delay, and jitter. These two mechanisms are the IP precedence setting in the IP header and the diff-serv code point (DSCP) defined in [3].

This document describes the various terms to be used in benchmarking devices that implement traffic control based on IP precedence or DSCP criteria. This document is narrowly focused, in that it describes only terms for measuring behavior of a device or a group of devices using one of these two mechanisms. End-to-end and service-level measurements are beyond the scope of this document.

This document introduces several new terms that will allow measurements to be taken during periods of congestion. New terminology is needed because most existing benchmarking terms assume the absence of congestion. For example, throughput is one of the most widely used measurements û- yet RFC 1242 defines throughput as a rate in the absence of loss. As a result, throughput is not a meaningful measurement where congestion exists.

Another key difference from existing benchmarking terminology is the definition of measurements as observed on egress as well as ingress of a device/system under test. Again, the existence of congestion requires the addition of egress measurements as well as those taken on ingress; without observing traffic leaving a device it is not possible to say whether traffic-control mechanisms effectively dealt with congestion.

The principal measurements introduced in this document are rate vectors, delay, and jitter, all of which can be observed with or without congestion of the DUT/SUT.

2. Existing definitions

RFC 1242 "Benchmarking Terminology for Network Interconnect Devices" and RFC 2285 "Benchmarking Terminology for LAN Switching Devices" should be consulted before attempting to make use of this document.

RFC 2474 "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers" section 2, contains discussions of a number of terms relevant to network-layer traffic control mechanisms and should also be consulted.

For the sake of clarity and continuity this RFC adopts the template for definitions set out in Section 2 of RFC 1242. Definitions are indexed and grouped together in sections for ease of reference.

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

3. Term definitions

3.1 Channel Capacity

Definition:
The maximum forwarding rate of a link or set of aggregated links at which none of the offered packets are dropped by the DUT/SUT.

Discussion:
Channel capacity measures the data rate at the egress interface(s) of the DUT/SUT. In contrast, throughput as defined in RFC 1242 measures the data rate is based on the ingress interface(s) of the DUT/SUT.

Ingress-based measurements do not account for congestion of the DUT/SUT. Channel capacity, as an egress measurement, does take congestion into account.

Understanding channel capacity is a necessary precursor to any

measurement involving congestion. Throughput numbers can be
higher than channel capacity because of queueing.

Measurement units:

N-octet packets per second

Issues:

See Also:
Throughput [1]

3.2 Classification

Definition:
Selection of packets based on the contents of packet header according to defined rules.

Discussion:
Packets can be selected based on the DS field or IP Precedence in the packet header. Classification can also be based on Multi-Field (MF) criteria such as IP Source and destination addresses, protocol and port number.

Classification determines the per-hop behaviors and traffic conditioning functions such as shaping and dropping that are to be applied to the packet.

Measurement units:

n/a

Issues:

See Also:
Rules

3.3 Codepoint Set

Definition:
The set of all DS Code-points or IP precedence values used during the test duration.

Discussion:
Describes all the code-point markings associated with packets that are input to the DUT/SUT. For each entry in the codepoint set, there are associated vectors describing the rate of traffic containing that particular DSCP or IP precedence value.

The treatment that a packet belonging to a particular code-point gets is subject to the DUT classifying packets to map to the correct PHB. Moreover, the forwarding treatment in general is also dependent on the complete set of offered vectors.

Measurement Units:
n/a

See Also:

3.4 Conforming

Definition:
Packets that lie within the bounds specified by a traffic profile.

Discussion:
Rules may be configured that allow a given traffic class to consume only X bit/s of channel capacity and no more. All additional packets are dropped. All packets that constitute the first X bits/s measured over a period of time specified by the traffic profile, are then said to be conforming whereas those exceeding the bound are non conforming.

In particular in a congestion scenario, some individual packets will be conforming and others will not.

Measurement units:

n/a

Issues:

See Also:
Expected Vector
Forwarding Vector
Offered Vector

3.5 Congestion

Definition:
A condition in which one or more egress interfaces are offered more packets than can be forwarded at any given instant.

Discussion:
This condition is a superset of the overload definition [2]. That definition assumes the overload is introduced strictly by

the tester on ingress of a DUT/SUT. That may or may not be the case here.

Another difference is that with multiple-DUT measurements, congestion may occur at multiple points. For example, multiple edge devices collectively may congest a core device. In contrast, overload [1] deals only with overload on ingress.

Ingress observations alone are not sufficient to cover all cases in which congestion may occur. A device with an infinite amount of memory could buffer an infinite amount of packets, and eventually forward all of them. However, these packets may or may not be forwarded during the test duration. Even though ingress interfaces accept all packets without loss, this hypothetical device may still be congested.

The definition presented here explicitly defines congestion as an event observable on egress interfaces. Regardless of internal architecture, any device that cannot forward packets on one or more egress interfaces is congested.

Measurement units:

n/a

Issues:

See Also:

3.6 Congestion Management

Definition:
An implementation of one or more per-hop-behaviors to avoid or minimize the condition of congestion.

Discussion:
Congestion management may seek either to control congestion or avoid it altogether. Such mechanisms classify packets based upon IP Precedence or DSCP settings in a packetÆs IP header.

Congestion avoidance mechanisms seek to prevent congestion before it actually occurs.

Congestion control mechanisms gives one or more service classes preferential treatment over other classes during periods of congestion.

Measurement units:

n/a

Issues:

See Also:

3.7 Delay

Definition:
The time interval starting when the last bit of the input IP packets reaches the input port of the DUT/SUT and ending when the last bit of the output IP packets is seen on the output port of the DUT/SUT.

Discussion:
Delay is measured the same regardless of the type of DUT/SUT. Latency [1] require some knowledge of whether the DUT/SUT is a "store and forward" or a "bit forwarding" device. The fact that a DUT/SUT's technology has a lower delay than another technology should be visible.

The measurement point at the end is more like the way an internet datagram is processed. An internet datagram is not passed up or down the stack unless it is complete. Completion occurs once the last bit of the IP packet has been received.

Delay can be run at any offered load. Recommend at or below the channel capacity for non-congested delay. For congested delay, run the offered load above the channel capacity.

Measurement units:

Seconds.

Issues:

See Also:

3.8 Expected Vector

Definition:
A vector describing the expected output rate of packets having a specific code-point. The value is dependent on the set of offered vectors and configuration of the DUT.

Discussion:
The DUT is configured in a certain way in order that service discrimination happens for behavior aggregates when a specific traffic mix consisting of multiple behavior aggregates is applied. This term attempts to capture the expected behavior, for which the device is configured, when subjected to a certain

offered load.

The actual algorithms or mechanism, that the DUT uses to achieve service discrimination, is not important in describing the expected vector.

Measurement units:
N-octets packets per second

See Also:
Forwarding Vector
Offered Vector
Codepoint Set

3.9 Flow

Definition:
A flow is a one or more of packets sharing a common intended pair of source and destination interfaces.

Discussion:
Packets are groups by the ingress and egress interfaces they use on a given DUT/SUT.

A flow can contain multiple source IP addresses and/or destination IP addresses. All packets in a flow must enter on the same ingress interface and exit on the same egress interface, and have some common network layer content.

Microflows [3] are a subset of flows. As defined in [3], microflows require application-to-application measurement. In contrast, flows use lower-layer classification criteria. Since this document focuses on network-layer classification criteria, we concentrate here on the use of network-layer identifiers in describing a flow. Flow identifiers also may reside at the data-link, transport, or application layers of the ISO model. However, identifiers other than those at the network layer are out of scope for this document.

A flow may contain a single code point/IP precedence value or may contain multiple values destined for a single egress interface. This is determined by the test methodology.

Measurement units:

n/a

Issues:

See Also:
  Microflow [3]

Streams

3.10 Forwarding Vector

Definition:
The number of packets per second for all packets containing a single DSCP (or IP precedence) that a device can be observed to successfully transmit to the correct destination interface in response to an offered vector.

Discussion:
Forwarding Vector is expressed as pair of numbers. Both the codepoint (or IP precedence) value AND the packets per second value combine to make a vector.

The forwarding vector represents packet rate based on their codepoint or IP precedence value. It is not necessary based on stream or flow. The forwarding vector may be expresses as ôper portö or ôof the DUT/SUTö.

Forwarding Vector is a per-hop measurement. The DUT/SUT may change the codepoint or IP precedence value for a multiple-hop measurement.

Measurement units:

N-octet packets per second

Issues:

See Also:
Codepoint Set
Expected Vector
Offered Vector.

3.11 Jitter

Definition:
Variation in a stream's delay.

Discussion:
Jitter is the absolute value of the difference between the delay measurement of two packets belonging to the same stream.

The jitter between two consecutive packets in a stream is reported as the "instantaneous jitter". Instantaneous jitter can be expressed as |D(i) û D(i+1)| where D equals the delay and i is the test sequence number. Packets lost are not counted in the jitter measurement.

Average Jitter is the average of the instantaneous jitter measured during the test duration.

Peak-to-peak jitter is the maximum delay minus the minimum delay of the packets forwarded by the DUT/SUT.

Measurement units:

Seconds (instantaneous)
Seconds P-P (peak to peak)
Seconds avg (average)

Issues:
Mean
Standard Deviation
Median
90th percentile
Inter Quartile Range

See Also:
Stream

3.12 Nonconforming

Definition:
Packets that lie outside the parameter bounds of a given traffic profile.

Discussion:
Rules may be configured for a given traffic class based on parameters, such as an upper bound on the rate of packet arrivals. All packets that lie outside the bounds specified by the traffic profile, measured over a period of time specified in the traffic profile, are said to be nonconforming.

Measurement units:

n/a

Issues:

See Also:
Conforming

3.13 Offered Vector

Definition:

A vector describing the rate at which packets having a specific code-point are offered to the DUT/SUT.

Discussion:
Offered loads across the different code-point classes, constituting a code-point set, determine the metrics associated with a specific code-point traffic class.

Measurement Units:
N-octets packets per second

Issues:
Packet size.

See Also:
Expected Vector
Forwarding Vector
Codepoint Set

3.14 Stream

Definition:
A group of packets tracked as a single entity by the traffic receiver. A stream shares a common content such as type (IP, UDP), frame size, or payload.

Discussion:
Streams are tracked by "sequence number" or "unique signature field" (RFC 2889). Streams define how individual packet's statistics are grouped together to form an intelligible summary.

Common stream groupings would be by egress interface, destination address, source address, DSCP, or IP precedence. A stream using sequence numbers can track the ordering of packets as they transverse the DUT/SUT.

Streams are not restricted to a pair of source and destination interfaces as long as all packets are tracked as a single entity. A mulitcast stream can be forward to multiple destination interfaces.

Measurement units:

n/a

Issues:

See Also:
  Flow
  MicroFlow [3]

Test sequence number

3.15 Tail dropping

Definition:
The condition in which a congested DUT/SUT discards newly arriving packets.

Discussion:
Every DUT/SUT has a finite amount of traffic it can forward, beyond which congestion occurs. Once the offered load crosses the congestion threshold, the device may discard any additional traffic that arrives until congestion clears.

Tail dropping is typically a function of offered load exceeding a DUT/SUTÆs buffer capacity, but other factors internal to the DUT/SUT may also come into play. In terms of what is externally observable, tail dropping can be said to occur only when offered load exceeds channel capacity. Since a DUT/SUT may buffer traffic on ingress, the actual threshold for tail dropping may be higher than channel capacity.

Measurement units:

n/a

Issues:
Some congestion management mechanisms seek to avoid tail dropping by discarding packets before offered load exceeds channel capacity. In the presence of such mechanisms, neither congestion nor tail dropping should occur.

See Also:
Channel capacity
Congestion

3.16 Test Sequence number

Definition:
A field in the IP payload portion of the packet that is used to verify the order of the packets on the egress of the DUT/SUT.

Discussion:
The traffic generator sets the sequence number value and the traffic receiver checks the value upon receipt of the packet. The traffic generator changes the value on each packet transmitted based on an algorithm agreed to by the traffic

receiver.

The traffic receiver keeps track of the sequence numbers on a per stream basis. In addition to number of received packets, the traffic receiver may also report number of in-sequence packets, number of out-sequence packets and number of duplicate packets.

The recommended algorithm to use to change the sequence number on sequential packets is an incrementing value.

Measurement units:

n/a

Issues:

See Also:
Stream

3.17 Unburdened Response

Definition:
A performance measure obtained when mechanisms used to support IP precedence and DiffServ are disabled.

Discussion:
Enabling Diffserv mechanisms such as scheduling algorithms may impose an additional processing overhead for packets, which may cause the aggregate response to suffer even when traffic belonging to only one class, the best effort class, is offered to the device. Comparisons with "unburdened performance" may thus be in order when obtaining metrics to ensure that enabling Diffserv mechanisms doesn't impose an excessive performance penalty.

Measurement units:

n/a

4. Security Considerations

Documents of this type do not directly effect the security of the Internet or of corporate networks as long as benchmarking is not performed on devices or systems connected to operating networks.

5. References

[1]   Bradner, S., Editor, "Benchmarking Terminology for Network

Interconnection Devices", RFC 1242, July 1991.

[2] Mandeville, R., "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998.

[3] K. Nichols, S. Blake, F. Baker, D. Black,"Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[4] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, W. Weiss, "An Architecture for Differentiated Services", RFC 2475, December 1998.

6. Author's Address

Jerry Perser
Spirent Communications
26750 Agoura Road
Calabasas, CA 91302
USA

Phone: + 1 818 676 2300
EMail: jerry.perser@spirentcom.com

David Newman
Network Test
31324 Via Colinas, Suite 113
Westlake Village, CA 91362
USA

Phone: + 1 818 889 0011, x10
EMail: dnewman@networktest.com

Sumit Khurana
Telcordia Technologies
445 South Street
Morristown, NJ 07960
USA

Phone: + 1 973 829 3170
EMail: sumit@research.telcordia.com

Shobha Erramilli
Telcordia Technologies

331 Newman Springs Road,
Navesink, NJ 07701
USA

Phone: + 1 732 758 5508
EMail: shobha@research.telcordia.com

Scott Poretsky
Avici Systems
101 Billerica AveùBuilding #6
N. Billerica, MA 01862
USA

Phone: + 1 978 964 2287
EMail: sporetsky@avici.com

7. Full Copyright Statement