Expiration Date: January 1997

Benchmarking Terminology for Local Area Switching Devices

<draft-ietf-bmwg-lanswitch-00.txt>

Abstract

The purpose of this draft is to extend the benchmarking terminology and methodology already defined for network interconnect devices in RFCs 1242 and 1944 by the Benchmarking Methodology Working Group (BMWG) of the Internet Engineering Task Force (IETF) to address the specific requirements of local area switches. Appendix A lists the tests and conditions that we believe should be included for specific cases and gives additional information about testing practices.

Although switches have clearly evolved from bridges, they have matured enough in the last few years to deserve special attention. Switches are seen as one of the principal sources of new bandwidth in the local area and are handling a significantly increasing proportion of network traffic. The multiplicity of products brought to market makes it desirable to define a set of benchmarks designed to provide reliable and comparable data to the user community with which to evaluate the performance characteristics of switching devices.

1. Introduction

The purpose of this draft is to discuss and define a number of terms and procedures for benchmarking switches. This draft covers local area devices which switch frames at the Media Access Control (MAC) layer. Its intention is to describe a benchmarking methodology which fully exercizes local area switching devices at the MAC layer. It defines tests for throughput, latency, address handling and filtering.

2. Term definitions

A previous document, "Benchmarking Terminology for Network Interconnect Devices" (RFC 1242), defined many of the terms that are used in this document. The terminology document should be consulted before attempting to make use of this document. A more recent document, "Benchmarking Methodology for Network Interconnect Devices" (RFC 1944), defined a number of test procedures which are directly applicable to switches. Since it discusses a number of terms relevant to benchmarking switches it should also be consulted. A number of new terms applicable to benchmarking switches are defined below using the format for definitions set out in Section 2 of RFC 1242. RFCs 1242 and 1944 already contain discussions of some of these terms.

2. 1. Reminder of RFC 1242 definition format

Term to be defined. (e.g., Latency)

Definition:
The specific definition for the term.

Discussion:
A brief discussion about the term, it's application and any restrictions on measurement procedures.

Measurement units:
The units used to report measurements of this term, if applicable.

Issues:
List of issues or conditions that effect this term.

See Also:
List of other terms that are relevant to the discussion of this term.

2.2. Unidirectional traffic

Definition:

Unidirectional traffic is made up of a single or multiple streams of frames forwarded in one direction only from one or more ports of a switching device designated as input ports to one or more other ports of the device designated as output ports.

Discussion:
This definition conforms to the discussion in section 16 of RFC 1944 on

multi-port testing which describes how unidirectional traffic can be offered to ports of a device to measure maximum rate of throughput.
With regard to benchmarking switching devices some additional applications of unidirectional traffic are to be considered:

- the measurement of the minimum inter-frame gap
- the detection of head of line blocking
- the measurement of throughput on ports when congestion control is activated
- the creation of many-to-one or one-to-many port overload
- the measurement of the aggressivity of the back-off algorithm in the case of CSMA/CD devices

A couple of these applications, head of line blocking and congestion control testing require unidirectional streams of traffic to be set up in a particular way between at least four ports with two streams running from one of the input ports to two output ports and a third stream running between the second input port and one of the output ports. These streams can be pictured as an inverted " Z " with input ports on the left and output ports on the right.
Many-to-one overload requires a minimum to two input and one output ports when all ports run at the same speed. When devices are equipped with ports running at different speeds the number of ports required to overload an output port or ports will vary.

Issues:
half duplex / full duplex

Measurement units:
n/a

See Also:
bidirectional traffic (2.3)
multidirectional traffic (2.4)

2.3. Bidirectional traffic

Definition:
Bidirectional traffic is made up of a single stream or multiple streams of frames forwarded in both directions between ports belonging to two distinct groups of ports on a switching device.

Discussion:
This definition conforms to the discussions in sections 14 and 16 of RFC **1944 on bidirectional traffic and multi-port testing.**
Bidirectional traffic MUST be offered when measuring the maximum rate of throughput on full duplex ports of a switching device.

It is not recommended to offer bidirectional traffic to measure maximum rates of throughput between isolated pairs of half duplex CSMA/CD ports since the capture effect may result in one of the ports transmitting for extended periods to the exclusion of the other port. The capture effect is generally considered to be an anomalous ramification of the truncated binary exponential back-off algorithm implemented in CSMA/CD devices.

Issues:
back-off

Measurement units:
n/a

See Also:
unidirectional traffic (2.2)
multidirectional traffic (2.4)

2.4. Multidirectional traffic

Definition:
Multidirectional traffic is made up of multiple streams of frames forwarded between all of the ports of a switching device.

Discussion:
This definition extends the discussions in sections 14 and 16 of RFC 1944 on bidirectional traffic and multi-port testing.
As with bidirectional multi-port tests, multidirectional traffic exercizes both the input and output sides of the ports of a switching device. But since ports are not divided into two groups every port forwards frames to and receives frames from every other port. The total number of individual unidirectional streams offered in a multidirectional test for n switched ports equals n x (n - 1). This compares with n x (n / 2) such streams in a bidirectional multi-port test. It should be noted however that bidirectional multiport tests create a greater load than multidirectional tests on backbone connections linking together two switching devices. Since none of the transmitted frames are forwarded locally all of the traffic is sent over the backbone. Backbone tests SHOULD use bidirectional multiport traffic.
Multidirectional traffic is inherently bursty since ports must interrupt transmission intermittently to receive frames. When offering such bursty traffic to a device under test a number of variables have to be defined. They include frame size, the number of frames within bursts as well as the interval between bursts. The terms burst size and inter-burst gap are defined in sections 2.6 and 2.7 below.
Bursty multidirectional traffic exercizes many of the component parts of a switching device simultaneously as they would be on a real network. It serves to determine the maximum throughput of a switching device when many of its componenet parts are working at once. Complementary tests may single out the performance characteristics of particular parts such as buffer size, backplane capacity, switching speed and the behavior of the media access controller . These tests are detailed in the methodology sections below.

Measurement units:
n/a

Issues:
half duplex / full duplex

See Also:
unidirectional traffic (2.2)
bidirectional traffic (2.3)
target rate / target load (2.6)

2.5 **Burst**

Definition:
A frame or a group of frames transmitted with the minimum inter-frame gap allowed by the media.

Discussion:
This definition follows from the discussion in section 21 of RFC 1944. It is useful to consider isolated frames as single frame bursts.

Measurement units:
n/a

Issues:

See Also:
burst size (2.6)

2.6 **Burst size**

Definition:
The number of frames in a burst.

Discussion:
Burst size can range from one to infinity. In unidirectional streams there is no theoretical limit to the burst length. Bursts in bidirectional and multidirectional streams of traffic are finite since ports interrupt transmission intermittantly to receive frames. In multidirectional networks bursts from several sources might be transmitted between ports at any one time. This makes it desirable to test devices for large burst sizes.

Measurement units:
number of N-octet frames

Issues:

See Also: burst (2.5)

2.7 **Inter-burst gap (IBG)**

Definition:
The interval between two bursts.

Discussion:
This definition conforms to the discussion in section 20 of RFC 1944 on bursty traffic.

Bidirectional and multidirectional streams of traffic are inherently bursty since ports share their time between receiving and transmitting frames. Assuming the number of frames per burst and frame length to be fixed, the value of the inter-burst gap will determine the rate of transmission. External sources offering bursty multidirectional traffic for a given frame size and burst size MUST adjust the inter-burst gap to achieve a specified rate of transmission.
When a burst contains a single frame inter-burst gap and inter-frame gap are equal.

Measurement units:
nanoseconds
microseconds
miliseconds
seconds

Issues:

See Also: burst size (2.6), load (2.8)

2.8 Load

Definition:
The amount of traffic per second going through the transmit and receive sides of a port.

Discussion:
Load can be expressed in a number of ways: bits per second, frames per second with the frame size specified or as a percentage of the maximum frame rate allowed by the media for a given frame size. For example, a port-to-port unidirectional stream of 7440 64-byte Ethernet frames per second offers a 50% load on the receive side of the input port and a 50% load on the transmit side of the output port given that the maximum line rate on an Ethernet is 14880 frames per second. In the case of bidirectional or multidirectional traffic port load is the sum of the frames received and transmitted on a port per second.
There is room for varying the balance between incoming and outgoing traffic when loading ports with bidirectional and multidirectional traffic. In the case of port-to-port bidirectional traffic a 100% load can be created by offering a n% load on the receive side of the input port and a (100 - n)% load on its transmit side. The output port will be offered the inverse load. Multidirectional traffic will be equally distributed over all ports under test when port receive and transmit sides are offered 50% loads. When benchmarking with balanced multidirectional loading ports under test MUST be offered an equally distributed load.
Target loads and actual loads may differ widely due to collisions on CSMA/CD links or the action of congestion control mechanisms. External sources of Ethernet traffic MUST implement the truncated binary exponential back-off algorithm when executing bidirectional and multidirectional performance tests to ensure that the external source of traffic is not accessing the medium illegally.

Frames which are not successfully transmitted by the external source of traffic to the device under test should be not counted as transmitted frames in performance benchmarks.

Measurement units:
bits per second
N-octets per second
(N-octets per second / media_maximum-octets per second) x 100

Issues:
token ring

2.9 Overload

Definition:
Loading a port or ports in excess of the maximum line rate allowed by the media.

Discussion:
Overloading can serve to test a device's buffer depth or congestion control mechanism. Unidirectional overloads require a minimum of two input and one output ports when all ports run at the same nominal speed. Balanced bidirectional and multidirectional overloading occur when the sum of the traffic offered to the input and output sides of all ports exceeds the maximum line rate allowed by the media by the same amount.

Measurement units:
N-octet frames per second

Issues: target load and mesured load

See Also:

2.10 Speed

Definition:
A measure of switching throughput which records the maximum number of frames that a switched port is capable of receiving and/or transmitting per second.

Discussion:
In multidirectional benchmarking it is important to record the speed at which switching devices are able to forward frames to their destination addresses. Speed can vary for a number of reasons such as head of line blocking, excessive collisions on CSMA/CD media, the action of congestion control mechanisms at high loads or the backplane capacity of the switching device. The rate of throughput on token rings is mostly a function of the media acces controllers.
The rate of throughput can be measured on the input as well as the output sides of a port. The rate of throughput measured on the output side of a port measures the rate at which a device forwards frames to their destinations. This rate MUST be reported as the rate of throughput. The aggregate rate of throughput can be skewed when a device drops frames since

the input port may receive at a much higher rate than it transmits.

Measurement units:
N-octet frames per second

Issues:

See Also:

2.11 Valid frame / invalid frame

Definition:
A frame which is forwarded to its proper destination port based on MAC address information is valid. A frame which is received on ports which do not correspond to the MAC address information is invalid.

Discussion:
When recording throughput statistics it is important to check that frames have been forwarded to their proper desinations. Invalid frames are generally unknown unicast frames which the device under test forwards or floods to all ports.

Measurement units:
N-octet valid frames per second

Issues:
Spanning tree BPDUs.

See Also:

2.11 Backpressure

Definition:
A jamming technique used by some switching devices to avoid frame loss when congestion on one or more of its ports occurs.

Discussion:
Some switches are designed to send jam signals, for example preamble bits, back to traffic sources when their transmit and/or receive buffers start to overfill. Such devices may incur no frame loss when ports are offered target loads in excess of 100% by external traffic sources. Jamming however affects traffic destined to congested as well as uncongested ports so it is important to measure the maximum speed at which a jamming port can forward frames to uncongested port destinations.

Measurement units:
N--octet frames per second between the jamming port and an uncongested destination port

Issues:

not explicitly described in standards

See Also:
forward pressure (2.12)

2.12 Forward pressure

Definition:
A technique which modifies the binary exponential backoff algorithm to avoid frame loss when congestion on one or more of its ports occurs.

Discussion:
Some switches avoid buffer overload by retransmitting buffered frames without waiting for the interval calculated by the normal operation of the backoff algorithm. It is important to measure how aggressive a switch's backoff algorithm is in both congested and uncongested states. Forward pressure is manifested by lower numbers of collisions when congestion on a port builds up.

Measurement units:
intervals in microseconds between transmission retries during 16 successive collisions.

Issues:
not explicitly described in standards

See also:
backpressure (2.11)

2.13 Head of line blocking

Definition:
A pathologocal state whereby a switch drops frames forwarded to an uncongested port whenever frames are forwarded from the same source port to a congested port.

Discussion:
It is important to verify that a switch does not propagate frame loss to ports which are not congested whenever overloading on one of its ports occurs.

Measurement units:
frame loss recorded on an uncongested port when receiving frames from a port which is also forwarding frames to a congested destination port.

Issues:
Input buffers

See Also:

2.14 Address handling

Definition:
The number of different destination MAC addresses which a switch can learn.

Discussion:

Users building networks will want to know how many nodes they can connect to a switch. This makes it necessary to verify the number of MAC addresses that can be assigned per port, per module and per chassis before a switch begins flooding frames.

Measurement units:
number of MAC addresses

Issues:

See Also:

2.15 Address learning speed

Definition:
The maximum rate at which a switch can learn MAC addresses before starting to flood frames.

Discussion:
Users may want to know how long it takes a switch to build up its address tables. This information may be useful for a user to have when considering how a network comes up after a crash.

Measurement units:
frames per second with each successive frame sent to the switch containing a different source address.

Issues:

See Also: address handling (2.14)

2.16 Filtering illegal frames

Definition:
Switches do not necessarily filter all types of illegal frames. Some switches, for example, do not store frames before forwarding them to their destination ports. These so-called cut-through switches forward frames after reading the destination and source address fields. They do not normally filter over-sized frames (jabbers) or verify the validity of the Frame Check Sequence field. Other illegal frame types are under-sized frames (runts), misaligned frames and frames followed by dribble bits.

Measurement units:
N-octet frames filtered or not filtered

Issues:

See Also:

2.17 Broadcast latency

Definition:
The time it takes a broadcast frame to go through a switching device and be forwarded to each destination port.

Discussion:
Since there is no standard way for switches to process broadcast frames, broadcast latency may not be the same on all receiving ports of a switching device. Broadcast latency SHOULD be determined on all receiving ports.

Measurement units:
The latency measurements SHOULD be bit oriented as described in 3.8 of RFC 1242 and reported for all connected receive ports.

Issues:

See Also:

3. Editor's Address

Robert Mandeville
ENL (European Network Laboratories)
email: bob.mandeville@eunet.fr
35, rue Beaubourg
75003 Paris
France
phone: +33 07 47 67 10
fax: + 33 1 42 78 36 71

Bob Mandeville
ENL
European Network Laboratories
office phone, fax and voice mail: +33 1 42 78 36 71
mobile phone: +33 07 47 67 10