Network Working Group
INTERNET-DRAFT
Expiration Date: August 1999

Hardev Soor
Debra Stopp
Ixia Communications
Ralph Daniels
Netcom Systems

February 1999

Methodology for IP Multicast Benchmarking

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

Status of this Memo

Abstract

The purpose of this draft is to describe methodology specific to the benchmarking of multicast IP forwarding devices. It builds upon the tenets set forth in RFC 1944, RFC 2432 and other IETF Benchmarking Methodology Working Group (BMWG) efforts. This document seeks to extend these efforts to the multicast paradigm.

The BMWG produces two major classes of documents: Benchmarking Terminology documents and Benchmarking Methodology documents. The Terminology documents present the benchmarks and other related terms. The Methodology documents define the procedures required to collect the benchmarks cited in the corresponding Terminology documents.

1. Introduction

This document defines a specific set of tests that vendors can use to measure and report the performance characteristics and forwarding capabilities of network devices that support IP multicast protocols. The results of these tests will provide the user comparable data from different vendors with which to evaluate these devices.

A previous document, " Terminology for IP Multicast Benchmarking" (RFC 2432), 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.

This methodology will focus on one source to many destinations, although many of the tests described may be extended to use multiple source to multiple destination IP multicast communication.

2. Key Words to Reflect Requirements

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. Test set up

Figure 1 shows a typical setup for an IP multicast test, with one source to multiple destinations, although this MAY be extended to multiple source to multiple destinations.

```
                                                   +----------------+
                           +------------+          |                |
        +--------+         |            |--------->| destination(1) |
        |        |         |            |          |                |
        | source |-------->|            |          +----------------+
        |        |         |            |          +----------------+
        +--------+         |   D U T    |--------->|                |
                           |            |          | destination(2) |
                           |            |          |                |
                           |            |          +----------------+
                           |            |               . . .
                           |            |          +----------------+
                           |            |          |                |
                           |            |--------->| destination(n) |
                           |            |          |                |
                           |            |          +----------------+
                           |            |
                           +------------+
                               Figure 1
```

Generally , the destination ports first join the desired number of multicast groups by sending IGMP Join Group messages to the DUT. To verify that all destination ports successfully joined the appropriate groups, the source port MUST transmit IP multicast frames destined for these groups.

In addition, all transmitted frames MUST contain a recognizable pattern that can be filtered on in order to ensure the receipt of only the frames that are involved in the test.

3.1 Test Considerations

3.1.1 IGMP Support

Each of the receiving ports should support and be able to test both IGMP version 1 and IGMP version 2.

Each receiving port should be able to respond to IGMP queries during the test.

Each receiving port should also send LEAVE (running IGMP version 2) after each test.

3.1.2 Group Addresses

The Class D Group address should be changed between tests. Many DUTs have memory or cache that is not cleared properly and can bias the results.

3.1.3 Frame Sizes

Each test should be run with different Multicast Frame Sizes.   The result to vary greatly.

3.1.4 TTL

The source frames should have a TTL value large enough to accommodate the DUT/SUT.

4. Forwarding and Throughput

This section contains the description of the tests that are related to the characterization of the packet forwarding of a DUT/SUT in a multicast environment. Some metrics extend the concept of throughput presented in RFC 1242. The notion of Forwarding Rate is cited in RFC 2285

4.1 Mixed Class Throughput

Definition

The maximum rate at which none of the offered frames, comprised from a unicast Class and a multicast Class, to be forwarded are dropped by the device across a fixed number of ports.

Procedure

Multicast and unicast traffic are mixed together in the same aggregated traffic stream in order to simulate the non-homogenous networking environment. While the multicast traffic is transmitted from one source to multiple destinations, the unicast traffic MAY be evenly distributed across the DUT architecture. In addition, the DUT SHOULD learn the appropriate unicast IP addresses, either by sending ARP frames from each unicast address, sending a RIP packet or by assigning static entries into the DUT address table.

The rates at which traffic is transmitted for both traffic classes MUST be set up in one of two ways:

a) A percentage of the bandwidth is allocated for each traffic class and frames for each class are transmitted at the rate equal to the allocated bandwidth. For example, 64 byte frames can be transmitted at a theoretical maximum rate of 148810 frames/second. If 80 percent of the bandwidth is allocated for unicast traffic and 20 percent for multicast traffic, then unicast traffic will be sent at a maximum rate of 119048 frames/second and the multicast traffic at a rate of 29762 frames/second.

b) Transmission rate is fixed for both traffic classes and a percentage of number of frames for each traffic class is specified. For example, if a fixed rate of 100% of theoretical maximum is desired, then 64 byte frames will be sent at 148810 frames/second for both unicast and multicast traffic. If 80 percent of the frames are to be unicast and 20 percent multicast, then for a duration of 10 seconds, 1190480 frames of unicast and 297620 frames of multicast will be sent. This fixed rate scenario actually over-subscribes the bandwidth, potentially causing congestion in the DUT.

The transmission of the frames MUST be set up so that they form a deterministic distribution while still maintaining the specified bandwidth and transmission rates. See Appendix A for a discussion on determining an even distribution.

Similar to the Frame loss rate test in RFC 1944, the first trial SHOULD be Run for the frame rate that corresponds to 100% of the maximum rate for the frame size on the input media. Repeat the procedure for the rate that corresponds to 90% of the maximum rate used and then for 80% of this rate. This sequence HOULD be continued (at reducing 10% intervals) until there are two successive trials in which no frames are lost. The maximum granularity of the trials MUST be 10% of the maximum rate, a finer granularity is encouraged.

Result

Transmit and Receive rates in frames per second for each source and destination port for both unicast and multicast traffic for each trial percent transmit rate. The result report SHOULD contain the number of frames transmitted and received per port per class type (unicast and multicast traffic), reported in number of frames and percent loss per port.

4.2 Scaled Group Forwarding Matrix

Definition:

A table that demonstrates Forwarding Rate as a function of tested multicast groups for a fixed number of tested DUT/SUT ports.

Procedure:

Multicast traffic is sent at a fixed percent of line rate with a fixed number of receive ports at a fixed frame length.

The receive ports will join an initial number of groups and the sender will transmit to the same groups after a certain delay (a few seconds).

Then the receive ports will join an incremental value of groups and the transmit port will send to all groups joined (initial plus incremental).

The receive ports will continue joining in the incremental fashion until a user defined maximum is reached.

Results:

For each group load the result WILL display frame rate, frames transmitted, total frames received, total frames loss, and percent loss. The frame loss per receive port per group SHOULD also be available.

4.3 Aggregated Multicast Throughput

Definition:

The maximum rate at which none of the offered frames to be forwarded through N destination interfaces of the same multicast group are dropped.

Procedure:

Multicast traffic is sent at a fixed percent of line rate with a fixed number of groups at a fixed frame length for a fixed duration of time.

The initial number of receive ports will join the group(s) and the sender will transmit to the same groups after a certain delay (a few seconds).

Then the an incremental number of receive ports will join the same groups and then the Multicast traffic is sent as stated.

The receive ports will continue to be added and the Multicast traffic sent until a user defined maximum number of ports is reached.

Results:

For each number of receive ports the result WILL display frame rate, frames transmitted, total frames received, total frames loss, and percent loss. The frame loss per receive port per group SHOULD also be available.

4.4 Encapsulation (Tunneling) Throughput

This sub-section provides the description of tests that help in obtaining throughput measurements when a DUT or a set of DUTs are acting as tunnel endpoints. The following Figure 2 presents the scenario for the tests.

```
      Client A        DUT A         Network            DUT B        Client B

                     ----------                     ----------
                     |        |      ------         |        |
      -------(a)  (b)|        |(c)  (      )   (d)|        |(e) (f)-------
      ||||||| -----> |        |---->(      )----->|        |-----> |||||||
      -------        |        |      ------         |        |       -------
                     |        |                     |        |
                     ----------                     ----------

                                 Figure 2
                                 --------
```

A tunnel is created between DUT A (the encapsulator) and DUT B (the decapsulator). Client A is acting as a source and Client B is the destination. Client B joins a multicast group (for example, 224.0.1.1) and it sends an IGMP Join message to DUT B to join that group. Client A now wants to transmit some traffic to Client B. It will send the multicast traffic to DUT A which encapsulates the multicast frames, sends it to DUT B which will decapsulate the same frames and forward them to Client B.

4.4.1 Encapsulation Throughput

Definition

The maximum rate at which frames offered a DUT are encapsulated and correctly forwarded by the DUT without loss.

Procedure

To test the forwarding rate of the DUT when it has to go through the process of encapsulation, a test port B is injected at the other end of DUT A (Figure B) that will receive the encapsulated frames and measure the throughput. Also, a test port A is used to generate multicast frames that will be passed through the tunnel.

The following is the test setup:

```
   Test port A      DUT A                  Test port B

                 ---------- (c')      (d')---------
                 |        |--------------->|       |
   -------(a)  (b)|        |               |       |
   ||||||| -----> |        |      ------   ---------
   -------        |        |(c)  ( N/W  )
                  |        |---->(      )
                 ----------      ------
```

Figure 3

In Figure 2, a tunnel is created with the local IP address of DUT A as the beginning of the tunnel (point c) and the IP address of DUT B as the end of the tunnel (point d). DUT B is assumed to have the tunneling protocol enabled so that the frames can be decapsulated. When the test port B is inserted in between the DUT A and DUT B (Figure 3), the endpoint of tunnel has to be re-configured to be directed to the test port B's IP address. For example, in Figure 3, point c' would be assigned as the beginning of the tunnel and point d' as the end of the tunnel. The test port B is acting as the end of the tunnel, and it does not have to support any tunneling protocol since the frames do not have to be decapsulated. Instead, the received encapsulated frames are used to calculate the throughput and other necessary measurements.

Result

Throughput in frames per second for each destination port. The results should also contain the number of frames transmitted and received per port.

4.4.2 Decapsulation Throughput

Definition

The maximum rate at which frames offered a DUT are decapsulated and correctly forwarded by the DUT without loss.

Procedure

The decapsulation process returns the tunneled unicast frames back to their multicast format. This test measures the throughput of the DUT when it has to perform the process of decapsulation, therefore, a test port C is used at the end of the tunnel to receive the decapsulated frames (Figure 4).

```
Test port A       DUT A       Test port B     DUT B        Test port C

               ----------                   ----------
               |        |                   |        |
-------(a)  (b)|        |(c)   ------    (d)|        |(e) (f)-------
||||||| -----> |        |----> |||||| ----->|        |-----> |||||||
-------        |        |      ------       |        |       -------
               |        |                   |        |
               ----------                   ----------
```

Figure 4

In Figure 4, the encapsulation process takes place in DUT A. This may effect the throughput of the DUT B. Therefore, two test ports should be used to separate the encapsulation and decapsulation processes. Client A is replaced with the test port A which will generate a multicast frame that will be encapsulated by DUT A. Another test port B is inserted between DUT A and DUT B that will receive the encapsulated frames and forward it to DUT B. Test port C will receive the decapsulated frames and measure the throughput.

Result

Throughput in frames per second for each destination port . The results should also contain the number of frames transmitted and received per port.

4.4.3 Re-encapsulation Throughput

Definition

The maximum rate at which frames of one encapsulated format offered a DUT are converted to another encapsulated format and correctly forwarded by the DUT without loss.

Procedure

Re-encapsulation takes place in DUT B after test port C has received the decapsulated frames. These decapsulated frames will be re-inserted with a new encapsulation frame and sent to test port B which will measure the throughput. See Figure 5.

```
     Test port A       DUT A       Test port B     DUT B       Test port C

                     ----------                  ----------
                     |        |                  |        |
      -------(a)  (b)|        |(c)   ------    (d)|        |(e) (f)-------
      ||||||| -----> |        |----> |||||| <---->|        |<----> |||||||
      -------        |        |      ------       |        |       -------
                     |        |                  |        |
                     ----------                  ----------
```

Figure 5

Result

Throughput in frames per second for each destination port. The results should also contain the number of frames transmitted and received per port.

5. Forwarding Latency

This section presents methodologies relating to the characterization of the forwarding latency of a DUT/SUT in a multicast environment. It extends the concept of latency characterization presented in RFC 1944.

5.1 Multicast Latency

Definition

The set of individual latencies from a single input port on the DUT or SUT to all tested ports belonging to the destination multicast group.

Procedure

According to RFC 1944, a tagged frame is sent half way through the transmission that contains a timestamp used for calculation of latency. In the multicast situation, a tagged frame is sent to all destinations for each multicast group and latency calculated on a per multicast group basis. Note that this test MUST be run using the transmission rate that is less than the multicast throughput of the DUT.

Result

The latency value for each multicast group address.

5.2 Min/Max/Average Multicast Latency

Definition:

The difference between the maximum latency measurement and the minimum latency measurement from the set of latencies produced by the Multicast Latency benchmark.

Procedure:

For the entire duration of the Latency test the smallest latency, the largest latency, the sum of latencies, and the number should be tracked per receive port.

The test can also increment bucket counters that represent a range latency range. This can be used to create a histogram. From the histogram, minimum, maximum, and average the test results can show the jitter.

Results:

For each port the results WILL display the number of frames, minimum latency, maximum latency, and the average latency. The results SHOULD also display the histogram of latencies.

6. Overhead

This section presents methodology relating to the characterization of the overhead delays associated with explicit operations found in multicast environments.

6.1 Group Join Delay

Definition:

The time duration it takes a DUT to start forwarding multicast packets from the time a successful IGMP group membership report has been issued to the DUT.

Procedure:

Traffic is sent on the source port at the same time as the IGMP JOIN Group message is transmitted from the destination ports. The join delay is the difference in time from when the IGMP Join is sent and the first frame is received.

One of the keys is to transmit at the fastest rate the DUT can handle multicast frames. This is to get the best resultion in the Join Delay. However, you do not want to transmit the frames to fast that frames are dropped by the DUT.

The test should be ran using both IGMPv1 JOIN and IGMPv2 JOIN.

Results:

The JOIN delay for each port.

6.2 Group Leave Delay

Definition

The time duration it takes a DUT to cease forwarding multicast packets after a corresponding IGMP "Leave Group" message has been successfully offered to the DUT.

Procedure

Traffic is sent on the source port at the same time as the IGMP Leave Group messages are transmitted from the destination ports. The frames on both the source and destination ports are sent with the timestamps inserted. The Group Leave Delay is the difference in the value of the timestamp A of the first IGMP Leave Group frame sent and the timestamp B of the last frame that is received on that destination port.

Group Leave delay = timestamp B - timestamp A

Result

Group Leave Delay values for each multicast group address on each destination port. Also, the number of frames transmitted and received, and percent loss may be displayed.

8. Capacity

This section offers terms relating to the identification of multicast group limits of a DUT/SUT.

8.1 Multicast Group Capacity

Definition:

The maximum number of multicast groups a SUT/DUT can support while maintaining the ability to forward multicast frames to all multicast groups registered to that SUT/DUT.

Procedure:

One or more receiving ports will join an initial number of groups. Then after a delay the source port will transmit to each group at a transmission rate that the DUT can handle.   If all frames sent are forwarded and received the receiving ports will join an incremental value of groups. Then after a a delay the source port will transmit to all groups at a transmission rate that the DUT can handle. If all frames sent are forwarded and received the receiving ports will continuing joining and testing until a frame is not forwarded nor received.

The group capacity resolution will be the incremental value. So the capacity could be greater then last capacity passed but less then the one that failed.

Once a capacity is determined the test should be re run with greater delays after the JOIN and a slower transmission rate. And the initial group level should be raised to about five less then the previous capacity and incremental value should be set to one.

Results:

The number of groups passed vs the number of groups failed. The results SHOULD give details when the frame fails to be forwarded about how many frames did and did not get forwarded. Which groups DID and DID NOT get forwarded.

9. Burdened Response

9.1 Burdened Response

Definition:

A measured response collected from a DUT/SUT in light of interacting, or potentially interacting, distinct stimuli.

Procedure:

TBD

Results:

TBD

9.2 Forwarding Burdened Multicast Latency

Definition

A multicast latency taken from a DUT/SUT in the presence of a traffic forwarding requirement.

Procedure

This test build upon the Multicast Latency as described in 1.5.1. This test measures latencies for a set of offered load values and reports the forwarding rates.

Result

A table of offered load (percent of maximum rate), forwarding rates and latencies for each multicast group involved.

9.3 Forwarding Burdened Join Delay

Definition:

A multicast Group Join Delay taken from a DUT in the presence of a traffic forwarding requirement.

Procedure:

TBD

Results:

TBD

Appendix A: Determining an even distribution

A.1 Scope Of This Appendix

This appendix discusses the suggested approach to configuring the deterministic distribution methodology for tests that involve both multicast and unicast traffic classes in an aggregated traffic stream. As such, this appendix MUST not be read as an amendment to the methodology described in the body of this document but as a guide to testing practice.

It is important to understand and fully define the distribution of frames among all multicast and unicast destinations. If the distribution is not well defined or understood, the throughput and forwarding metrics are not meaningful.

In a homogeneous environment, a large, single burst of multicast frames may be followed by a large burst of unicast frames. This is a very different distribution than that of a non-homogeneous environment, where the multicast and unicast frames are intermingled throughout the entire transmission.

The recommended distribution is that of the non-homogeneous environment because it more closely represents a real-world scenario. The distribution is modeled by calculating the number of multicast frames per destination port as a burst, then calculating the number of unicast frames to transmit as a percentage of the total frames transmitted. The overall effect of the distribution is small bursts of multicast frames intermingled with small bursts of unicast frames.

Example

```
Frame size = 64
Duration of test = 10 seconds
Transmission rate = 100% of maximum rate
Mapping for unicast traffic:    Port 1 to Port 2
                                Port 3 to port 4
Mapping for multicast traffic:  Port 1 to Ports 2,3,4
Number of Multicast group addresses per destination port = 3
Multicast groups joined by Port 2: 224.0.1.1
                                   224.0.1.2
```

```
                                   224,0.1.3
      Multicast groups joined by Port 3: 224.0.1.4
                                   224.0.1.5
                                   224,0.1.6
      Multicast groups joined by Port 4: 224.0.1.7
                                   224.0.1.8
                                   224,0.1.9

      Percentage of Unicast frames = 20
      Percentage of Multicast frames = 80
      Total number of frames to be transmitted = 148810 fps * 10 sec
                                         = 1488100 frames
      Number of unicast frames = 20/100 * 1488100 = 297620 frames
      Number of multicast frames = 80/100 * 1488100 = 1190480 frames

      Unicast burst size = 20 * 9 = 180
      Multicast burst size = 80 * 9 = 720
      Loop counter = 1488100 / 900 = 1653.4444 (round it off to 1653)

      Therefore, the actual number of frames that will be transmitted:
        Unicast frames = 1653 * 180 = 297540 frames
        Multicast frames = 1653 * 720 = 1190160 frames

      The following pattern will be established:

      UUUMMMMMMMMMMMMUUUMMMMMMMMMMMMUUUMMMMMMMMMMMMUUUMMMMMMMMMMMM

      where     U represents 60 Unicast frames (UUU = 180 frames)
                M represents 60 Multicast frames (MMMMMMMMMMMM = 720 frames)
```

3. Security Considerations.

As this document is solely for the purpose of providing metric methodology and describes neither a protocol nor a protocol's implementation, there are no security considerations associated with this document.

5. References

[Br91] Bradner, S., "Benchmarking Terminology for Network Interconnection Devices", RFC 1242, July 1991.

[Br96] Bradner, S., and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 1944, May 1996.

[Br97] Bradner, S. "Use of Keywords in RFCs to Reflect Requirement Levels, RFC 2119, March 1997

[Du98] Dubray, K., "Terminology for IP Multicast Benchmarking", RFC 2432, October 1998.

[Hu95] Huitema, C. "Routing in the Internet." Prentice-Hall, 1995.

[Ka98] Kosiur, D., "IP Multicasting: the Complete Guide to Interactive Corporate Networks", John Wiley & Sons, Inc, 1998.

[Ma98] Mandeville, R., "Benchmarking Terminology for LAN Switching Devices", RFC 2285, February 1998.

[Mt98] Maufer, T. "Deploying IP Multicast in the Enterprise." Prentice-Hall, 1998.

[Se98] Semeria, C. and Maufer, T. "Introduction to IP Multicast Routing." http://www.3com.com/nsc/501303.html 3Com Corp., 1998.

6. Author's Address

Hardev Soor
Ixia Communications
4505 Las Virgenes Road, Suite 209
Calabasas, CA 91302
USA

Phone: 818 871 1800
EMail: hardev@ixiacom.com

Debra Stopp
Ixia Communications
4505 Las Virgenes Road, Suite 209
Calabasas, CA 91302
USA

Phone: 818 871 1800
EMail: debby@ixiacom.com

Ralph Daniels
Netcom Systems
948 Loop Road
Clayton, NC 27520
USA

Phone: 919 550 9475
EMail: Ralph_Daniels@NetcomSystems.com