INTERNET-DRAFT
Expires in six months

Benchmarking Methodology for Network Interconnect Devices

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

Status of this Document

This document is an Internet-Draft. Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet- Drafts as reference material or to cite them other than as ``work in progress.''

To learn the current status of any Internet-Draft, please check the ``1id-abstracts.txt'' listing contained in the Internet-Drafts Shadow Directories on ds.internic.net (US East Coast), nic.nordu.net (Europe), ftp.isi.edu (US West Coast), or munnari.oz.au (Pacific Rim).

Distribution of this document is unlimited. Please send comments to bmwg@harvard.edu or to the editor.

Abstract

This document discusses and defines a number of tests that may be used to describe the performance characteristics of a network interconnecting device. In addition to defining the tests this document also describes specific formats for reporting the results of the tests. Appendix A lists the tests and conditions that we believe should be included for specific cases and gives additional information about testing practices. Appendix B is a reference listing of maximum frame rates to be used with specific frame sizes on various media and Appendix C gives some examples of frame formats to be used in testing.

1. **Introduction**

Vendors often engage in "specsmanship" in an attempt to give their products a better position in the marketplace. This often involves "smoke & mirrors" to confuse the potential users of the products. This document and follow up memos attempt to define a specific set of tests that vendors can use to measure and report the performance characteristics of network devices. The results of these tests will provide the user comparable data from different vendors with which to evaluate these devices.

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.

2. Real world

In producing this document the authors attempted to keep in mind the requirement that apparatus to perform the described tests must actually be built. We do not know of "off the shelf" equipment available to implement all of the tests but it is our opinion that such equipment can be constructed.

3. Tests to be run

There are a number of tests described in this document. Not all of the tests apply to all types of devices. It is expected that a vendor will perform all of the tests that apply to a specific type of product. The authors understand that it will take a considerable period of time to perform all of the recommended tests under all of the recommended conditions. We believe that the results are worth the effort. Appendix A lists the tests and conditions that we believe should be included for specific cases.

4. Evaluating the results

Performing all of the recommended tests will result in a great deal of data. Much of this data will not apply to the evaluation of the devices under each circumstance. For example, the rate at which a router forwards IPX frames will be of little use in selecting a router for an environment that does not (and will not) support that protocol. Evaluating even that data which is relevant to a particular network installation will require experience which may not be readily available.

5. Requirements

In this document, the words that are used to define the significance of each particular requirement are capitalized. These words are:

* "MUST"

This word or the adjective "REQUIRED" means that the item is an absolute requirement of the specification.

* "SHOULD"

This word or the adjective "RECOMMENDED" means that there may exist valid reasons in particular circumstances to ignore this item, but the full implications should be understood and the case carefully weighed before choosing a different course.

* "MAY"

This word or the adjective "OPTIONAL" means that this item is truly optional. One vendor may choose to include the item because a particular marketplace requires it or because it enhances the product, for example; another vendor may omit the same item.

An implementation is not compliant if it fails to satisfy one or more of the MUST requirements for the protocols it implements. An implementation that satisfies all the MUST and all the SHOULD requirements for its protocols is said to be "unconditionally compliant"; one that satisfies all the MUST requirements but not all the SHOULD requirements for its protocols is said to be "conditionally compliant".

6. Device set up

Before starting to perform the tests, the device to be tested MUST be configured following the instructions provided to the user. Specifically, it is expected that all of the supported protocols will be configured and enabled during this set up (See Appendix A). It is expected that all of the tests will be run without changing the configuration or setup of the device in any way other than that required to do the specific test. For example, it is not acceptable to change the size of frame handling buffers between tests of frame handling rates or to disable all but one transport protocol when testing the throughput of that protocol. It is necessary to modify the configuration when starting a test to determine the effect of filters on throughput, but the only change MUST be to enable the specific filter. The device set up SHOULD include the normally recommended routing update intervals and keep alive frequency. The specific version of the software and the exact device configuration, including what device functions are disabled, used during the tests SHOULD be included as part of the report of the results.

7. Frame formats

The formats of the test frames to use for TCP/IP over Ethernet are shown in Appendix C: Test Frame Formats. It is expected that these exact frame formats will be used in the tests described in this document for this protocol/media combination and that these frames will be used as a template for testing other protocol/media combinations. The specific formats that are used to define the test frames for a particular test series MUST be included in the report of the results.

8. Frame sizes

All of the described tests SHOULD be performed at a number of frame sizes. Specifically, the sizes SHOULD include the maximum and minimum legitimate sizes for the protocol under test on the media under test and enough sizes in between to be able to get a full characterization of the device performance.

Except where noted, it is expected that at least five frame sizes will be tested for each test condition.

Theoretically the minimum size UDP Echo request frame would consist of an IP header (minimum length 20 octets), a UDP header (8 octets) and whatever MAC level header is required by the media in use. The theoretical maximum frame size is determined by the size of the length field in the IP header. In almost all cases the actual maximum and minimum sizes are determined by the limitations of the media.

In theory it would be ideal to distribute the frame sizes in a way that would evenly distribute the theoretical frame rates. These recommendations incorporate this theory but specify frame sizes which are easy to understand and remember. In addition, many of the same frame sizes are specified on each of the media types to allow for easy performance comparisons.

The inclusion of an unrealistically small frame size on some of the media types (i.e. with little or no space for data) is to help characterize the per-frame processing overhead of the network connection device.

8.1 Frame sizes to be used on Ethernet

64, 128, 256, 512, 1024, 1280, 1518

These sizes include the maximum and minimum frame sizes permitted by the Ethernet standard and a selection of sizes between these extremes with a finer granularity for the smaller frame sizes and higher frame rates.

8.2 Frame sizes to be used on 4Mb and 16Mb token ring

54, 64, 128, 256, 1024, 1518, 2048, 4472

The frame size recommendations for token ring assume that there is no RIF field in the frames of routed protocols. A RIF field would be present in any direct source route bridge performance test. The minimum size frame for UDP on token ring is 54 octets. The maximum size of 4472 octets is recommended for 16Mb token ring instead of the theoretical size of 17.9Kb because of the size limitations imposed by many token ring interfaces. The reminder of the sizes are selected to permit direct comparisons with other types of media. An IP (i.e. not UDP) frame may be used in addition if a higher data rate is desired, in which case the minimum frame size is 46 octets.

8.3 Frame sizes to be used on FDDI

54, 64, 128, 256, 1024, 1518, 2048, 4472

The minimum size frame for UDP on FDDI is 53 octets, the minimum size of **54 is recommended to allow direct comparison to token ring performance.** The maximum size of 4472 is recommended instead of the theoretical maximum size of 4500 octets to permit the same type of comparison. An IP (i.e. not UDP) frame may be used in addition if a higher data rate is desired, in which case the minimum frame size is 45 octets.

9. Verifying received frames

The test equipment SHOULD discard any frames received during a test run that are not actual forwarded test frames. For example, keep-alive and routing update frames SHOULD NOT be included in the count of received frames. In any case, the test equipment SHOULD verify the length of the received frames and check that they match the expected length.

Preferably, the test equipment SHOULD include sequence numbers in the transmitted frames and check for these numbers on the received frames. If this is done, the reported results SHOULD include in addition to the number of frames dropped, the number of frames that were received out of order, the number of duplicate frames received and the number of gaps in the received frame numbering sequence. This functionality is required for some of the described tests.

10. Modifiers

It might be useful to know the device performance under a number of conditions; some of these conditions are noted below.   It is expected that the reported results will include as many of these conditions as the test equipment is able to generate. The suite of tests SHOULD be first run without any modifying conditions and then repeated under each of the conditions separately. To preserve the ability to compare the results of these tests any frames that are required to generate the modifying conditions (management queries for example) will be included in the same data stream as the normal test frames in place of one of the test frames and not be supplied to the device on a separate network port.

10.1   Broadcast frames

In most router designs special processing is required when frames addressed to the hardware broadcast address are received. In bridges (or in bridge mode on routers) these broadcast frames must be flooded to a number of ports. The stream of test frames SHOULD be augmented with 1% frames addressed to the hardware broadcast address. The specific frames that should be used are included in the test frame format document. The broadcast frames SHOULD be evenly distributed throughout the data stream, for example, every 100th frame.

It is understood that a level of broadcast frames of 1% is much higher than many networks experience but, as in drug toxicity evaluations, the higher level is required to be able to gage the effect which would otherwise often fall within the normal variability of the system performance. Due to design factors some test equipment will not be able to generate a level of alternate frames this low. In these cases it is expected that the percentage would be as small as the equipment can provide and that the actual level be described in the report of the test results.

10.2   Management frames

Most data networks now make use of management protocols such as SNMP. In many environments there can be a number of management stations sending queries to the same device at the same time.

The stream of test frames SHOULD be augmented with one management query as the first frame sent each second during the duration of the trial. The result of the query must fit into one response frame. The response frame SHOULD be verified by the test equipment. One example of the specific query frame that should be used is shown in Appendix C.

10.3 Routing update frames

The processing of dynamic routing protocol updates could have a significant impact on the ability of a router to forward data frames. The stream of test frames SHOULD be augmented with one routing update frame transmitted as the first frame transmitted during the trial. Routing update frames SHOULD be sent at the rate specified in Appendix C for the specific routing protocol being used in the test. Two routing update frames are defined in Appendix C for the TCP/IP over Ethernet example. The routing frames are designed to change the routing to a number of networks that are not involved in the forwarding of the test data. The first frame sets the routing table state to "A", the second one changes the state to "B". The frames MUST be alternated during the trial.

The test SHOULD verify that the routing update was processed by the device under test.

10.4 Filters

Filters are added to routers and bridges to selectively inhibit the forwarding of frames that would normally be forwarded. This is usually done to implement security controls on the data that is accepted between one area and another. Different products have different capabilities to implement filters.

The device SHOULD be first configured to add one filter condition and the tests performed. This filter SHOULD permit the forwarding of the test data stream. In routers this filter SHOULD be of the form:

```
        forward input_protocol_address to output_protocol_address
```

In bridges the filter SHOULD be of the form:

```
        forward destination_hardware_address
```

The device SHOULD be then reconfigured to implement a total of 25 filters. The first 24 of these filters SHOULD be of the form:

```
        block input_protocol_address to output_protocol_address
```

The 24 input and output protocol addresses SHOULD not be any that are represented in the test data stream. The last filter SHOULD permit the forwarding of the test data stream. By "first" and "last" we mean to ensure that in the second case, 25 conditions must be checked before the data frames will match the conditions that permit the forwarding of the frame.

The exact filters configuration command lines used SHOULD be included with the report of the results.

10.4.1 Filter Addresses

Two sets of filter addresses are required, one for the single filter case and one for the 25 filter case.

The single filter case should permit traffic from IP address 198.18.1.2 to IP address 198.19.65.2 and deny all other traffic.

The 25 filter case should follow the following sequence.

```
        allow aa.ba.1.1 to aa.ba.100.1
        allow aa.ba.2.2 to aa.ba.101.2
        allow aa.ba.3.3 to aa.ba.103.3
                 ...
        allow aa.ba.12.12 to aa.ba.112.12
        allow aa.bc.1.2 to aa.bc.65.1
        allow aa.ba.13.13 to aa.ba.113.13
        allow aa.ba.14.14 to aa.ba.114.14
                 ...
        allow aa.ba.24.24 to aa.ba.124.24
        deny all else
```

All previous filter conditions should be cleared from the router before this sequence is entered. The sequence is selected to test to see if the router sorts the filter conditions or accepts them in the order that they were entered. Both of these procedures will result in a greater reduction in performance than will some form of hash coding.

11.   Protocol addresses

It is easier to implement these tests using a single logical stream of data, with one source protocol address and one destination protocol address, and for some conditions like the filters described above, a practical requirement. Networks in the real world are not limited to single streams of data. The test suite SHOULD be first run with a single protocol (or hardware for bridge tests) source and destination address pair. The tests SHOULD then be repeated with using a random destination address. While testing routers the addresses SHOULD be random over a range of 256 networks and random over the full MAC range for bridges. The specific address ranges to use for IP are shown in Appendix C.

12.   Route Set Up

It is not expected that all of the routing information necessary to forward the test stream, especially in the multiple address case, will be manually set up. At the start of each trial a routing update MUST be sent to the device. This routing update MUST include all of the network addresses that will be required for the trial. All of the addresses SHOULD resolve to the same "next-hop" and it is expected that this will be the address of the receiving side of the test equipment. This routing update will have to be repeated at the interval required by the routing protocol being used. An example of the format and repetition interval of the update frames is given in Appendix C.

13. Bidirectional traffic

Normal network activity is not all in a single direction. To test the bidirectional performance of a device, the test series SHOULD be run with the same data rate being offered from each direction. The sum of the data rates should not exceed the theoretical limit for the media.

14. Single stream path

The full suite of tests SHOULD be run along with whatever modifier conditions that are relevant using a single input and output network port on the device. If the internal design of the device has multiple distinct pathways, for example, multiple interface cards each with multiple network ports, then all possible types of pathways SHOULD be tested separately.

15. Multi-port

Many current router and bridge products provide many network ports in the same device. In performing these tests first half of the ports are designated as "input ports" and half are designated as "output ports". These ports SHOULD be evenly distributed across the device architecture. For example if a device has two interface cards each of which has four ports, two ports on each interface card are designated as input and two are designated as output.

The specified tests are run using the same data rate being offered to each of the input ports. The addresses in the input data streams SHOULD be set so that a frame will be directed to each of the output ports in sequence. The stream offered to input one SHOULD consist of a series of frames (one destined to each of the output ports), as SHOULD the frame stream offered to input two. The same configuration MAY be used to perform a bidirectional multi-stream test. In this case all of the ports are considered both input and output ports and each data stream MUST consist of frames addressed to all of the other ports.

16. Multiple protocols

This document does not address the issue of testing the effects of a mixed protocol environment other than to suggest that if such tests are wanted then frames SHOULD be distributed between all of the test protocols. The distribution MAY approximate the conditions on the network in which the device would be used.

17. Multiple frame sizes

This document does not address the issue of testing the effects of a mixed frame size environment other than to suggest that if such tests are wanted then frames SHOULD be distributed between all of the listed sizes for the protocol under test. The distribution MAY approximate the conditions on the network in which the device would be used.

18. Testing performance beyond a single device.

In the performance testing of a single device, the paradigm can be described as applying some input to a device under test and monitoring the output. The results of which can be used to form a basis of characterization of that device under those test conditions.

This model is useful when the test input and output are homogenous (e.g., 64-byte IP, 802.3 frames into the device under test; 64 IP, 802.3 frames out), or the method of test can distinguish between dissimilar input/output. (E.g., 1518 byte, IP, 802.3 frames in; 576 byte, fragmented IP, X.25 frames out.)

By extending the single device test model, reasonable benchmarks regarding multiple devices or heterogeneous environments may be collected. In this extension, the single device under test is replaced by a system of interconnected network devices. This test methodology would support the benchmarking of a variety of device/media/service/protocol combinations. For example, a configuration for a LAN-to-WAN-to-LAN test might be:

(1) 802.3-> device 1 -> X.25 @ 64kbps -> device 2 -> 802.3

Or a mixed LAN configuration might be:

(2) 802.3 -> device 1 -> FDDI -> device 2 -> FDDI -> device 3 -> 802.3

In both examples 1 and 2, end-to-end benchmarks of each system could be empirically ascertained. Other behavior may be characterized through the use of intermediate devices. In example 2, the configuration may be used to give an indication of the FDDI to FDDI capability exhibited by device 2.

Because multiple devices are treated as a single system, there are limitations to this methodology. For instance, this methodology may yield an aggregate benchmark for a tested system. That benchmark alone, however, may not necessarily reflect asymmetries in behavior between the devices, latencies introduce by other apparatus (e.g., CSUs/DSUs, switches), etc.

Further, care must be used when comparing benchmarks of different systems by ensuring that the devices' features/configuration of the tested systems have the appropriate common denominators to allow comparison.

The maximum frame rate that should be used when testing WAN connections SHOULD be greater than the listed theoretical maximum rate for the frame size on that speed connection. The higher rate for WAN tests is to compensate for the fact that some vendors employ various forms of header compression. See Appendix A.

19.   Maximum frame rate

The maximum frame rate that should be used when testing LAN connections SHOULD be the listed theoretical maximum rate for the frame size on the media. A list of maximum frame rates for LAN connections is included in Appendix B.

20. Bursty traffic

It is convenient to measure the device performance under steady state load but this is an unrealistic way to gage the functioning of a device since actual network traffic normally consists of bursts of frames. Some of the tests described below SHOULD be performed with both steady state traffic and with traffic consisting of repeated bursts of frames. The frames within a burst are transmitted with the minimum legitimate inter-frame gap.

The objective of the test is to determine the minimum interval between bursts which the device under test can process with no frame loss. During each test the number of frames in each burst is held constant and the inter-burst interval varied. Tests SHOULD be run with burst sizes of 16, 64, 256 and 1024 frames.

21. Frames per token

Although it is possible to configure some token ring and FDDI interfaces to transmit more than one frame each time that the token is received, most of the network devices currently available transmit only one frame per token. These tests SHOULD first be performed while transmitting only one frame per token.

Some current high-performance workstation servers do transmit more than one frame per token on FDDI to maximize throughput. Since this may be a common feature in future workstations and servers, interconnect devices with FDDI interfaces SHOULD be tested with 1, 4, 8, and 16 frames per token. The reported frame rate SHOULD be the average rate of frame transmission over the total trial period.

22. Trial description

A particular test consists of multiple trials. Each trial returns one piece of information, for example the loss rate at a particular input frame rate. Each trial consists of a number of phases:

a) If the test device is a router, send the routing update to the "input" port and pause two seconds to be sure that the routing has settled.

b) Send the "learning frames" to the "output" port and wait 2 seconds to be sure that the learning has settled. Bridge learning frames are frames with source addresses that are the same as the destination addresses used by the test frames. Learning frames for other protocols are used to prime the address resolution tables in the device. The formats of the learning frame that should be used are shown in the Test Frame Formats document.

c) Run the test trial.

d) Wait for two sec for any residual frames to be received.

e) Wait for at least five seconds for the device to restabilize.

23. Trial duration

The aim of these tests is to determine the rate continuously supportable by the device. The actual duration of the test trials must be a compromise between this aim and the duration of the benchmarking test suite. The duration of the test portion of each trial SHOULD be at least 60 seconds. The tests that involve some form of "binary search", for example the throughput test, to determine the exact result MAY use a shorter trial duration to minimize the length of the search procedure, but it is expected that the final determination will be made with full length trials.

24 Address resolution

The test device SHOULD be able to respond to address resolution requests sent by the device under test wherever the protocol requires such a process.

25 Benchmarking tests:

Note: The notation "type of data stream" refers to the above modifications to a frame stream with a constant inter-frame gap, for example, the addition of traffic filters to the configuration of the device under test.

25.1 Throughput

Objective:
To determine the device throughput as defined in RFC 1242.

Procedure:
Send a specific number of frames at a specific rate through the device and then count the frames that are transmitted by the device. If the count of offered frames is equal to the count of received frames, the rate of the offered stream is raised and the test rerun. If fewer frames are received than were transmitted, the rate of the offered stream is reduced and the test is rerun.

The throughput is the fastest rate at which the count of test frames transmitted is equal to the number of test frames sent.

Reporting format:
The results of the throughput test SHOULD be reported in the form of a graph. If it is, the x coordinate SHOULD be the frame size, the y coordinate SHOULD be the frame rate. There SHOULD be at least two lines on the graph. There SHOULD be one line showing the theoretical frame rate for the media at the various frame sizes. The second line SHOULD be the plot of the test results. Additional lines MAY be used on the graph to report the results for each type of data stream tested. Text accompanying the graph SHOULD indicate the protocol, data stream format, and type of media used in the tests.

We assume that if a single value is desired for advertising purposes the vendor will select the rate for the minimum frame size for the media. If this is done then the figure MUST be expressed in frames per second. The rate MAY also be expressed in bits (or bytes) per second if the vendor so desires. The statement of performance MUST include a/ the measured maximum frame rate, b/ the size of the frame used, c/ the theoretical limit of the media for that frame size, and d/ the type of protocol used in the test. Even if a single value is used as part of the advertising copy, the full table of results SHOULD be included in the product data sheet.

25.2 Latency

Objective:
To determine the latency as defined in RFC 1242.

Procedure:
First determine the throughput for device at each of the listed frame sizes.

Send a stream of frames at a particular frame size through the device at the determined throughput rate to a specific destination. The stream SHOULD be at least 120 seconds in duration. An identifying tag SHOULD be included in one frame after 60 seconds with the type of tag being implementation dependent. The time at which this frame is fully transmitted is recorded, i.e. the last bit has been transmitted (timestamp A). The receiver logic in the test equipment MUST be able to recognize the tag information in the frame stream and record the time at which the entire tagged frame was received (timestamp B).

The latency is timestamp B minus timestamp A minus the transit time for a frame of the tested size on the tested media. This calculation may result in a negative value for those devices that begin to transmit the output frame before the entire input frame has been received.

The test MUST be repeated at least 20 times with the reported value being the average of the recorded values.

This test SHOULD be performed with the test frame addressed to the same destination as the rest of the data stream and also with each of the test frames addressed to a new destination network.

Reporting format:
The latency results SHOULD be reported in the format of a table with a row for each of the tested frame sizes. There SHOULD be columns for the frame size, the rate at which the latency test was run for that frame size, for the media types tested, and for the resultant latency values for each type of data stream tested.

25.3 Frame loss rate

Objective:
To determine the frame loss rate, as defined in RFC 1242, of a device throughout the entire range of input data rates and frame sizes.

Procedure:
Send a specific number of frames at a specific rate through the device to be tested and count the frames that are transmitted by the device. The frame loss rate at each point is calculated using the following equation:

$$((\text{input\_count} - \text{output\_count}) * 100) / \text{input\_count}$$

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 SHOULD 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.

Reporting format:
The results of the frame loss rate test SHOULD be plotted as a graph. If this is done then the X axis MUST be the input frame rate as a percent of the theoretical rate for the media at the specific frame size. The Y axis MUST be the percent loss at the particular input rate. The left end of the X axis and the bottom of the Y axis MUST be 0 percent; the right end of the X axis and the top of the Y axis MUST be **100 percent.** Multiple lines on the graph MAY used to report the frame loss rate for different frame sizes, protocols, and types of data streams.

Note: See section 18 for the maximum frame rates that SHOULD be used.

25.4 Back-to-back frames

Objective:
To characterize the ability of a device to process back-to-back frames as defined in RFC 1242.

Procedure:
Send a burst of frames with minimum inter-frame gaps to the device and count the number of frames forwarded by the device. If the count of transmitted frames is equal to the number of frames forwarded the length of the burst is increased and the test is rerun. If the number of forwarded frames is less than the number transmitted, the length of the burst is reduced and the test is rerun.

The back-to-back value is the number of frames in the longest burst that the device will handle without the loss of any frames.

The trial length MUST be at least 2 seconds and SHOULD be repeated at least 50 times with the average of the recorded values being reported.

Reporting format:
The back-to-back results SHOULD be reported in the format of a table

with a row for each of the tested frame sizes. There SHOULD be columns for the frame size and for the resultant average frame count for each type of data stream tested. The standard deviation for each measurement MAY also be reported.

25.5 System recovery

Objective:
To characterize the speed at which a device recovers from an overload condition.

Procedure:
First determine the throughput for a device at each of the listed frame sizes.

Send a stream of frames at a rate 110% of the recorded throughput rate or the maximum rate for the media, whichever is lower, for at least 60 seconds. At Timestamp A reduce the frame rate to 50% of the above rate and record the time of the last frame lost (Timestamp B). The system recovery time is determined by subtracting Timestamp A from Timestamp B. The test SHOULD be repeated a number of times and the average of the recorded values being reported.

Reporting format:
The system recovery results SHOULD be reported in the format of a table with a row for each of the tested frame sizes. There SHOULD be columns for the frame size, the frame rate used as the throughput rate for each type of data stream tested, and for the measured recovery time for each type of data stream tested.

25.6 Reset

Objective:
To characterize the speed at which a device recovers from a device or software reset.

Procedure:
First determine the throughput for the device for the minimum frame size on the media used in the testing.

Send a continuous stream of frames at the determined throughput rate for the minimum sized frames. Cause a reset in the device. Monitor the output until frames begin to be forwarded and record the time that the last frame (Timestamp A) of the initial stream and the first frame of the new stream (Timestamp B) are received.

A power interruption reset test is performed as above except that the power to the device should be interrupted for 10 seconds in place of causing a reset.

This test SHOULD only be run using frames addressed to networks directly connected to the device under test so that there is no requirement to delay until a routing update is received.

The reset value is obtained by subtracting Timestamp A from Timestamp B.

Hardware and software resets, as well as a power interruption SHOULD be tested.

Reporting format:
The reset value SHOULD be reported in a simple set of statements, one for each reset type.

26. Security Considerations

Security issues are not addressed in this document.

27. Editor's Address

Scott Bradner
Holyoke Center   Phone +1 617 495-3864
Harvard University   Fax +1 617 495-0914
Cambridge, MA 02138   Email: sob@harvard.edu

Jim McQuaid
Wandel & Goltermann Technologies, Inc   Phone +1 919 941-4730
P. O. Box 13585   **Fax: +1 919 941-5751**
Research Triangle Park, NC 27709   Email: mcquaid@wg.com

Appendix A: Testing Considerations

A.1 Scope Of This Appendix

This appendix discusses certain issues in the benchmarking methodology where experience or judgement may play a role in the tests selected to be run or in the approach to constructing the test with a particular device. 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.

1. **Typical testing practice has been to enable all protocols to be** tested and conduct all testing with no further configuration of protocols, even though a given set of trials may exercise only one protocol at a time. This minimizes the opportunities to "tune" a device under test for a single protocol.

2. **The least common denominator of the available filter functions** should be used to ensure that there is a basis for comparison between vendors. Because of product differences, those conducting and evaluating tests must make a judgement about this issue.

3. **Architectural considerations may need to be considered.** For example, first perform the tests with the stream going between ports on the same interface card and the repeat the tests with the stream going into a port on one interface card and out of a port on a second interface card. There will almost always be a best case and worst case configuration for a given device under test architecture.

4. **Testing done using traffic streams consisting of mixed protocols** has not shown much difference between testing with individual protocols. That is, if protocol A testing and protocol B testing give two different performance results, mixed protocol testing appears to give a result which is the average of the two.

5. **Wide Area Network (WAN) performance may be tested by setting up** two identical devices connected by the appropriate short-haul versions of the WAN modems. Performance is then measured between a LAN interface on one device to a LAN interface on the other device.

The maximum frame rate to be used for LAN-WAN-LAN configurations is a judgement that can be based on known characteristics of the overall system including compression effects, fragmentation, and gross link speeds. Practice suggests that the rate should be at least 110% of the slowest link speed. Substantive issues of testing compression itself are beyond the scope of this document.

Appendix B: Maximum frame rates reference

(Provided by Roger Beeman)

Ethernet Size	Ethernet	16Mb Token Ring	FDDI
(bytes)	(pps)	(pps)	(pps)
64	14880	24691	152439
128	8445	13793	85616
256	4528	7326	45620
512	2349	3780	23585
768	1586	2547	15903
1024	1197	1921	11996
1280	961	1542	9630
1518	812	1302	8138

Ethernet size

Preamble	64 bits
Frame	8 x N bits
Gap	96 bits

16Mb Token Ring size

SD	8 bits

AC	8 bits
FC	8 bits
DA	48 bits
SA	48 bits
RI	48 bits (06 30 00 12 00 30)
SNAP	
DSAP	8 bits
SSAP	8 bits
Control	8 bits
Vendor	24 bits
Type	16 bits
Data	8 x (N - 18) bits
FCS	32 bits
ED	8 bits
FS	8 bits

No accounting for token or idles between packets (theoretical minimums hard to pin down)

FDDI size

Preamble	64 bits
SD	8 bits
FC	8 bits
DA	48 bits
SA	48 bits
SNAP	
DSAP	8 bits
SSAP	8 bits
Control	8 bits
Vendor	24 bits
Type	16 bits
Data	8 x (N - 18) bits
FCS	32 bits
ED	4 bits
FS	12 bits

No accounting for token or idles between packets (theoretical minimums hard to pin down)

Appendix C: Test Frame Formats

This appendix defines the frame formats that may be used with these tests. It also includes protocol specific parameters for TCP/IP over Ethernet to be used with the tests as an example.

C.1.   Introduction

The general logic used in the selection of the parameters and the design of the frame formats is explained for each case within the TCP/IP section. The same logic has been used in the other sections. Comments are used in these sections only if there is a protocol specific feature to be explained. Parameters and frame formats for additional protocols can be defined by the reader by using the same logic.

C.2.   TCP/IP Information

The following section deals with the TCP/IP protocol suite.

C.2.1   Frame Type.

An application level datagram echo request is used for the test data frame in the protocols that support such a function. A datagram protocol is used to minimize the chance that a router might expect a specific session initialization sequence, as might be the case for a reliable stream protocol. A specific defined protocol is used because some routers verify the protocol field and refuse to forward unknown protocols.

For TCP/IP a UDP Echo Request is used.

C.2.2   Protocol Addresses

Two sets of addresses must be defined: first the addresses assigned to the router ports, and second the address that are to be used in the frames themselves and in the routing updates.

The following specific network addresses are have been assigned to the BMWG by the NIC for this purpose. This assignment was made to minimize the chance of conflict in case a testing device were to be accidentally connected to part of the Internet.

C.2.2.1 Router port protocol addresses

Half of the ports on a multi-port router are referred to as "input" ports and the other half as "output" ports even though some of the tests use all ports both as input and output. A contiguous series of IP Class C network addresses from 198.18.1.0 to 198.18.64.0 have been assigned for use on the "input" ports. A second series from 198.19.1.0 to 198.19.64.0 **have been assigned for use on the "output" ports. In all** cases the router port is node 1 on the appropriate network. For example, a two port device would have an IP address of 198.18.1.1 on one port and 198.19.1.1 on the other port.

Some of the tests described in the methodology memo make use of an SNMP management connection to the device under test. The management access address for the device is assumed to be the first of the "input" ports (198.18.1.1).

C.2.2.2 Frame addresses

Some of the described tests assume adjacent network routing (the reboot time test for example). The IP address used in the test frame is that of node 2 on the appropriate Class C network. (198.19.1.2 for example)

If the test involves non-adjacent network routing the phantom routers are located at node 10 of each of the appropriate Class C networks. A series of Class C network addresses from 198.18.65.0 to 198.18.254.0 has been assigned for use as the networks accessible through the phantom routers on the "input" side of device under test. The series of Class C networks from 198.19.65.0 to 198.19.254.0 have been assigned to be used as the networks visible through the phantom routers on the "output" side of the device under test.

C.2.3 Routing Update Frequency

The update interval for each routing protocol is may have to be determined by the specifications of the individual protocol. For IP RIP, Cisco IGRP and for OSPF a routing update frame or frames should precede each stream of test frames by 5 seconds. This frequency is sufficient for trial durations of up to 60 seconds. Routing updates must be mixed with the stream of test frames if longer trial periods are selected. The frequency of updates should be taken from the following table.

IP-RIP	30 sec
IGRP	90 sec
OSPF	90 sec

C.2.4 Frame Formats - detailed discussion

C.2.4.1 Learning Frame

In most protocols a procedure is used to determine the mapping between the protocol node address and the MAC address. The Address Resolution Protocol (ARP) is used to perform this function in TCP/IP. No such procedure is required in XNS or IPX because the MAC address is used as the protocol node address.

In the ideal case the tester would be able to respond to ARP requests from the device under test. In cases where this is not possible an ARP request should be sent to the router's "output" port. This request should be seen as coming from the immediate destination of the test frame stream. (i.e. the phantom router or the end node if adjacent network routing is being used.) It is assumed that the router will cache the MAC address of the requesting device. The ARP request should be sent 5 seconds before the test frame stream starts in each trial. Trial lengths of longer than 50 seconds may require that the router be configured for an extended ARP timeout.

C.2.4.2 Routing Update Frame

If the test does not involve adjacent net routing the tester must supply proper routing information using a routing update. A single routing update is used before each trial on each "destination" port (see section C.24). This update includes the network addresses that are reachable through a phantom router on the network attached to the port. For a full mesh test, one destination network address is present in the routing update for each of the "input" ports. The test stream on each "input" port consists of a repeating sequence of frames, one to each of the "output" ports.

C.2.4.3 Management Query Frame

The management overhead test uses SNMP to query a set of variables that should be present in all devices that support SNMP. The variables are read by an NMS at the appropriate intervals. The list of variables to retrieve follow:

```
        sysUpTime
        ifInOctets
        ifOutOctets
        ifInUcastPkts
        ifOutUcastPkts
```

C.2.4.4 Test Frames

The test frame is an UDP Echo Request with enough data to fill out the required frame size. The data should not be all bits off or all bits on since these patters can cause a "bit stuffing" process to be used to maintain clock synchronization on WAN links. This process will result in a longer frame than was intended.

C.2.4.5 Frame Formats - TCP/IP on Ethernet

Each of the frames below are described for the 1st pair of device ports, i.e. "input" port #1 and "output" port #1. Addresses must be changed if the frame is to be used for other ports.

C.2.6.1 Learning Frame

```
        ARP Request on Ethernet

        -- DATAGRAM HEADER
        offset  data (hex)              description
        00      FF FF FF FF FF FF       dest MAC address
                                                send to broadcast address
        06      xx xx xx xx xx xx       set to source MAC address
        12      08 06                   ARP type
        14      00 01                   hardware type
                                                Ethernet = 1
        16      08 00                   protocol type
                                                IP = 800
        18      06                      hardware address length
                                                48 bits on Ethernet
        19      04                      protocol address length
                                                4 octets for IP
        20      00 01                   opcode
                                                request = 1
        22      xx xx xx xx xx xx       source MAC address
        28      xx xx xx xx             source IP address
        32      FF FF FF FF FF FF       requesting DUT's MAC address
        38      xx xx xx xx             DUT's IP address
```

C.2.6.2 Routing Update Frame

```
        -- DATAGRAM HEADER
        offset  date                    description
        00      FF FF FF FF FF FF       dest MAC address is broadcast
        06      xx xx xx xx xx xx       source hardware address
        12      08 00                   type

        -- IP HEADER
        14      45                      IP version - 4,
                                                header length
                                                (4 byte units) - 5
        15      00                      service field
        16      00 EE                   total length
        18      00 00                   ID
        20      40 00                   flags (3 bits)
                                                4 (do not fragment),fragment
offset-0
        22      0A                      TTL
        23      11                      protocol - 17 (UDP)
        24      C4 8D                   header checksum
        26      xx xx xx xx             source IP address
        30      xx xx xx                destination IP address
        33      FF                      host part = FF for broadcast

        -- UDP HEADER
        34      02 08                   source port
                                                208 = RIP
        36      02 08                   destination port
                                                208 = RIP
        38      00 DA                   UDP message length
        40      00 00                   UDP checksum

        -- RIP packet
        42      02                      command = response
        43      01                      version = 1
        44      00 00                   0

        -- net 1
        46      00 02                   family = IP
        48      00 00                   0
        50      xx xx xx                net 1 IP address
        53      00                      net not node
        54      00 00 00 00             0
        58      00 00 00 00             0
        62      00 00 00 07             metric 7
        -- net 2
        66      00 02                   family = IP
        68      00 00                   0
        70      xx xx xx                net 2 IP address
        73      00                      net not node
        74      00 00 00 00             0
        78      00 00 00 00             0
        82      00 00 00 07             metric 7

        -- net 3
        86      00 02                   family = IP
        88      00 00                   0
        90      xx xx xx                net 3 IP address
        93      00                      net not node
        94      00 00 00 00             0
        98      00 00 00 00             0
        102     00 00 00 07             metric 7

        -- net 4
        106     00 02                   family = IP
        108     00 00                   0
        110     xx xx xx                net 4 IP address
        113     00                      net not node
        114     00 00 00 00             0
        118     00 00 00 00             0
        122     00 00 00 07             metric 7

        -- net 5
        126     00 02                   family = IP
        128     00 00                   0
        130     00                      net 5 IP address
        133     00                      net not node
        134     00 00 00 00             0
        138     00 00 00 00             0
        142     00 00 00 07             metric 7

        -- net 6
        146     00 02                   family = IP
        148     00 00                   0
        150     xx xx xx                net 6 IP address
        153     00                      net not node
        154     00 00 00 00             0
        158     00 00 00 00             0
        162     00 00 00 07             metric 7
```

C.2.4.6 Management Query Frame

To be defined.

C.2.6.4 Test Frames

```
     UDP echo request on Ethernet

        -- DATAGRAM HEADER
        offset  data                    description
        00      xx xx xx xx xx xx       set to dest MAC address
        06      xx xx xx xx xx xx       set to source MAC address
        12      08 00                   type

        -- IP HEADER
        14      45                      IP version - 4
                                                header length 5 4 byte units
        15      00                      TOS
        16      00 2E                   total length*
        18      00 00                   ID
        20      00 00                   flags (3 bits) - 0
                                                fragment offset-0
        22      0A                      TTL
        23      11                      protocol - 17 (UDP)
        24      C4 8D                   header checksum*
        26      xx xx xx xx             set to source IP address**
        30      xx xx xx xx             set to destination IP address**

        -- UDP HEADER
        34      C0 20                   source port
        36      00 07                   destination port
                                                07 = Echo
        38      00 1A                   UDP message length*
        40      00 00                   UDP checksum

        -- UDP DATA
        42      00 01 02 03 04 05 06 07    some data***
        50      08 09 0A 0B 0C 0D 0E 0F
```

* - change for different length frames

** - change for different logical streams

*** - fill remainder of frame with incrementing octets, repeated if required by frame length

Values to be used in Total Length and UDP message length fields:

frame size	total length	UDP message length
64	**00 2E**	00 1A
128	**00 6E**	00 5A
256	**00 EE**	00 9A
512	**01 EE**	01 9A
768	**02 EE**	02 9A
1024	**03 EE**	03 9A
1280	**04 EE**	04 9A
1518	**05 DC**	05 C8

Benchmarking Methodology Working Group Scott Bradner
Internet Draft - November 1994   Harvard University
Jim McQuaid
Wandel & Goltermann Technologies, Inc