Bit-error Rate Testing with the Lanforge

20 July 2006

Revision History

Rev	Date	Summary	Author
1	19 July 2006	First version of document.	amarburg@mbari.org
2	20 July 2006	Rearranged the hardware and performed another throughput test	amarburg@mbari.org

Table of Contents

Summary

After some initial testing with the Lanforge traffic simulator,

- We found no significant data error problems within the system while operating a 2.4kW inductive load on science port 8 (July 18, 2006)
- The LanForge unit reaches a natural limit at approximately 500Mbits/sec from a single gigabit interface. The DCS system was able to pass this amount of data over the node-to-shore connection without incident (July 19, 2006)

Setup

Hardware was configured as follows:

Lanforge Port	NIC	Role	Plugged to
Eth0	Intel Pro/1000	Topside Ops	Port 2/0/12 on topside mra91-02
Eth1	Intel Pro/1000	Node Ops	Port 1/0/11 on node router mra01-01
Eth2	Intel Pro/1000	(mgmnt)	Port 2/0/11 on topside mra91-02
Eth3	Intel Pro/1000	Science 8	End of test pigtail
Eth4	Digital 21143	Science 2	Port 1/0/2 on node router mra01-01
Eth5	Digital 21143	Science 3	Port 1/0/3 on node router mra01-01
Eth6	Digital 21143	Science 4	Port 1/0/4 on node router mra01-01
Eth7	Digital 21143	Science 5	Port 1/0/5 on node router mra01-01

The Intel Pro/1000 interfaces are capable of gigabit speeds, the Digital 21143 interfaces are only capable of fast (100 Mbit/sec) speeds. All interfaces were left at default configuration, except port eth1 (node ops) was run with an MTU setting of 5000 and a TX queue length of 10,000 (the MTU value was chosen to match the max packet size below, the TX queue length was picked arbitrarily). [An attempt was made to run the topside ops interface eth0 with the same settings but performance *decreased* noticeably. Still working on that…]

Cross connections were established between the topside ops port (eth0) and each of the node side ports. Each connection was loaded with one or more of three traffic streams:

- TCP 56kbit/sec with 300 byte packets in both directions (symmetrically).
- TCP 50Mbit/sec with 300-5000 bytes packets node-to-shore, 56kbit/sec with 300 byte packets shore-to-node.
- UDP 50Mbit/sec with 300-5000 bytes packets node-to-shore, 56kbit/sec with 300 byte packets shore-to-node.

All streams were bidirectional, passing data in both directions. The listed speeds are desired values which the LanForge attempts to produce. It's unclear what role the packet sizes play – presumably the LanForge software will send out as large a packet as possible to get the target throughput.

The 56kbps/300byte symmetric stream is the LanForge default for new connections, and seems a reasonable representation for a moderate interactive connection (i.e. perhaps analogous to the PMACS server to NPC connection – though I haven't measured actual traffic on that connection). 50 Mbps is the fastest rate the LanForge will generate for either a UDP or TCP stream and was selected to simulate a very large data stream – perhaps live video or acoustic data. The allowable maximum packet size was increased to 5000 bytes as it seemed to improve path throughput. The data stream in the opposite direction (shore-to-node) was kept at 56kbps to simulate low-bandwidth control signals.

Only the node side operations network connection (eth1) had its metric transfer unit (MTU) set to a value other than the default of 1500. As such, the actual packet size on any of the other interfaces may actually be capped at 1500 bytes (still investigating this).

TCP is a stateful connection algorithm which will attempt to retransmit packets if they are lost or garbled. UDP is a stateless connection and is unable to recover from any sort of packet loss. As such, problems with TCP data streams will tend not to show up as lost or corrupted data, but as high numbers of retransmissions and a commensurate drop in total throughput. UDP streams will happily lose data.

The LanForge software interface provides real-time information on the current state of each connection, most notably the current throughput in each direction and the packet loss rate. For TCP connections, the packet loss rate should be 0% (or 100% if not working at all). As noted in the LanForge manual non-zero values may come about if a packet is sitting in a kernel buffer, waiting for the LanForge to process it. UDP connections, however, can and will lose packets.

The LanForge software also provides information on each of the physical interfaces on the LanForge computer, notably transmit collisions and receive errors, which may come from malformed packets, packet corruption, etc.

It is also possible to interrogate similar interface information from each of the Cisco 3750 routers in the system.

Unfortunately, it is difficult to get precise statistical information from the LanForge (i.e. time history of throughput, etc). Still working on that problem.

Dry tests with thruster (18 July 2006)

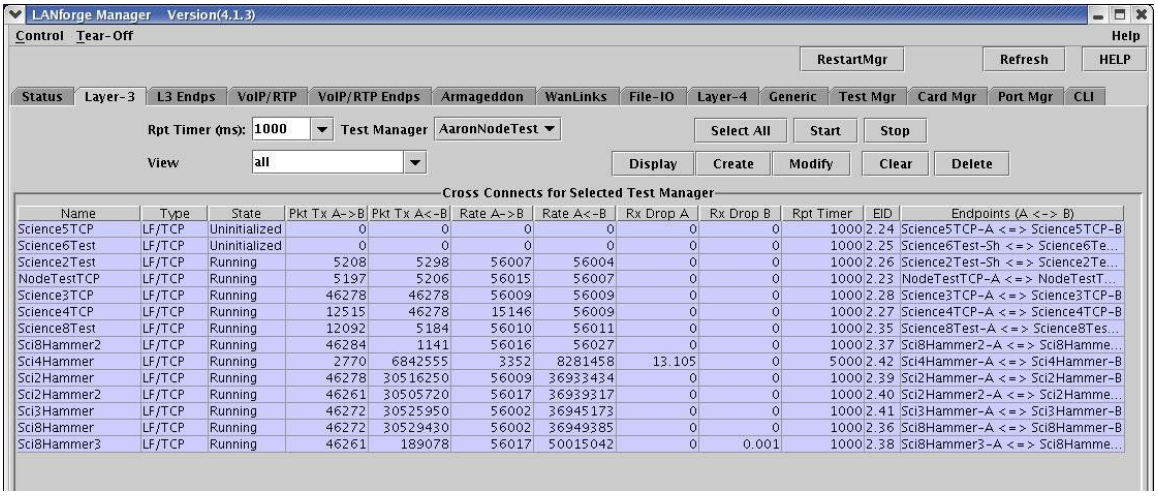
The first round of testing was scheduled to determine the effect (if any) of moderate inductive loads on Ethernet communications. The thruster/dynamometer setup was attached to the 375V connection on the test pigtail on science port 8. System voltage was lowered to 325V (the thruster motor is only designed to operate to a max of 350V). Over the course of a half hour the power draw by the motor was stepped up from 0W to 2.4kW, staying at 2.4kW for approximately 5 minutes.

The Lanforge was configured as follows (all tests are TCP):

- Node operations: one 56kbps symmetric stream
- Science 2: one 56kbps symmetric, two 50Mbps/56kbps asymmetric
- Science 3 and 4: one 56kbps symmetric and one 50Mbps/56kbps asymmetric each
- Science 8: two 56kbps symmetric, two 50Mbbps/56kbps asymmetric

for a total theoretical throughput of 250.3Mbps from the node and 560kbps to the node.

An image from the LanForge test manager at the conclusion of the power test can be seen below:



Name	Type	State	Pkt Tx A->B	Pkt Tx A<-B	Rate A->B	Rate A<-B	Rx Drop A	Rx Drop B	Rpt Timer	EID	Endpoints (A <-> B)
Science5TCP	LF/TCP	Uninitialized	0	0	0	0	0	0	1000	2.24	Science5TCP-A <=> Science5TCP-B
Science6Test	LF/TCP	Uninitialized	0	0	0	0	0	0	1000	2.25	Science6Test-Sh <=> Science6Te...
Science2Test	LF/TCP	Running	5208	5298	56007	56004	0	0	1000	2.26	Science2Test-Sh <=> Science2Te...
NodeTestTCP	LF/TCP	Running	5197	5206	56015	56007	0	0	1000	2.23	NodeTestTCP-A <=> NodeTestT...
Science3TCP	LF/TCP	Running	46278	46278	56009	56009	0	0	1000	2.28	Science3TCP-A <=> Science3TCP-B
Science4TCP	LF/TCP	Running	12515	46278	15146	56009	0	0	1000	2.27	Science4TCP-A <=> Science4TCP-B
Science8Test	LF/TCP	Running	12092	5184	56010	56011	0	0	1000	2.35	Science8Test-A <=> Science8Tes...
Sci8Hammer2	LF/TCP	Running	46284	1141	56016	56027	0	0	1000	2.37	Sci8Hammer2-A <=> Sci8Hamme...
Sci4Hammer	LF/TCP	Running	2770	6842555	3352	8281458	13.105	0	5000	2.42	Sci4Hammer-A <=> Sci4Hammer-B
Sci2Hammer	LF/TCP	Running	46278	30516250	56009	36933434	0	0	1000	2.39	Sci2Hammer-A <=> Sci2Hammer-B
Sci2Hammer2	LF/TCP	Running	46261	30505720	56017	36939317	0	0	1000	2.40	Sci2Hammer2-A <=> Sci2Hamme...
Sci3Hammer	LF/TCP	Running	46272	30525950	56002	36945173	0	0	1000	2.41	Sci3Hammer-A <=> Sci3Hammer-B
Sci8Hammer	LF/TCP	Running	46272	30529430	56002	36949385	0	0	1000	2.36	Sci8Hammer-A <=> Sci8Hammer-B
Sci8Hammer3	LF/TCP	Running	46261	189078	56017	50015042	0	0.001	1000	2.38	Sci8Hammer3-A <=> Sci8Hamme...

The tests labeled “Test” or “TCP” are 56kbps symmetric connections. Those labeled “hammer” are 50Mbps/56kbps asymmetric (except “Sci8Hammer2” which is 56k symmetric, a naming error). In each case, side “A” is the shore and side “B” is the node. The columns labeled “Pkt Tx” give the number of packets transmitted. The “Rx Drop”

columns give a percentage of those packets which are not properly received at the other interface. The "Rate" columns give the actual throughput in bits-per-second in each direction averaged over the previous 30 seconds.

With the exception of those connections involving science port 4, all of the shore-to-node connection maintained their specified rates (56k in every case). The 56kpbs node-to-shore streams were also maintained, although the system struggled somewhat to maintain 50MBit on all of the asymmetric connections.

There were a large number of receive errors involving science port 4 which significantly retarded its throughput. As this test was done with non-flight equipment, very little time was taken to diagnose the problem.

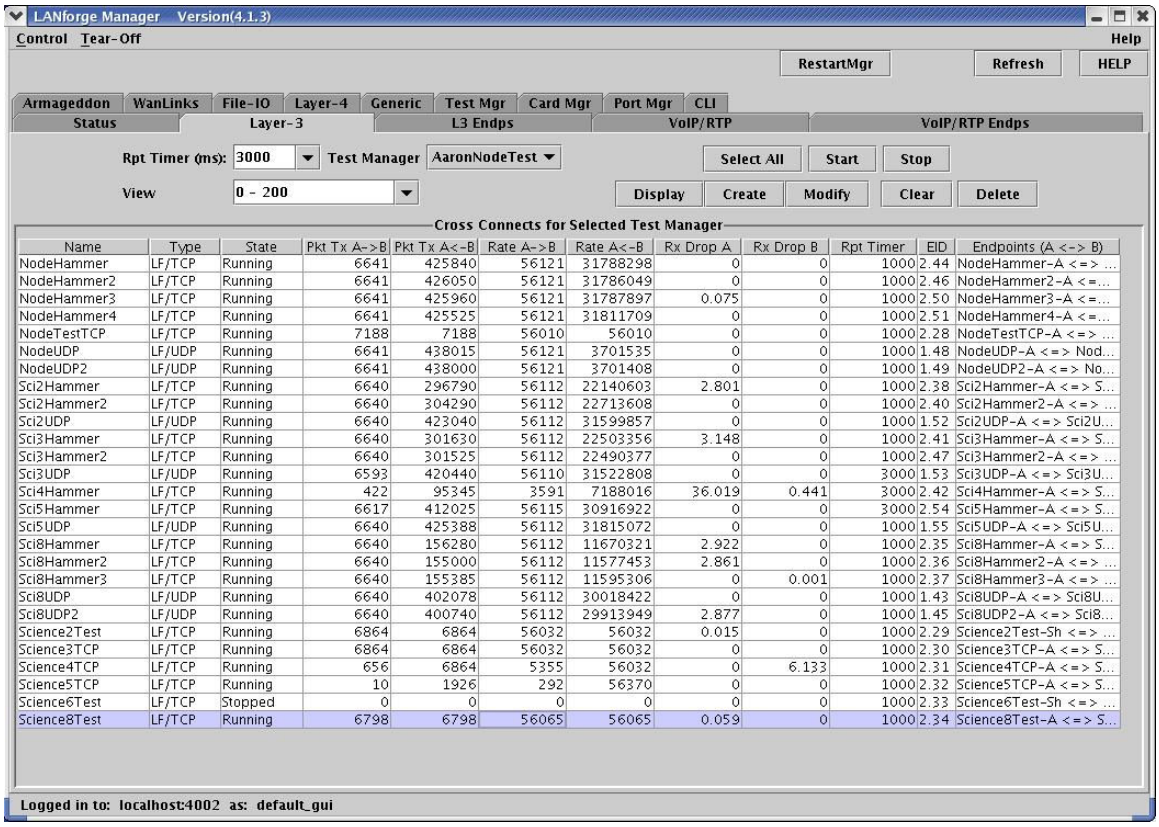
There was no measurable increase in error rate with the motor running. Each of the science ports other than port 4 showed between 0 and 2 CRC errors from the approximately half hour of running. Aggregate node-to-shore bandwidth was approximately 200 Mbits per second.

Max throughput testing (19 July 2006)

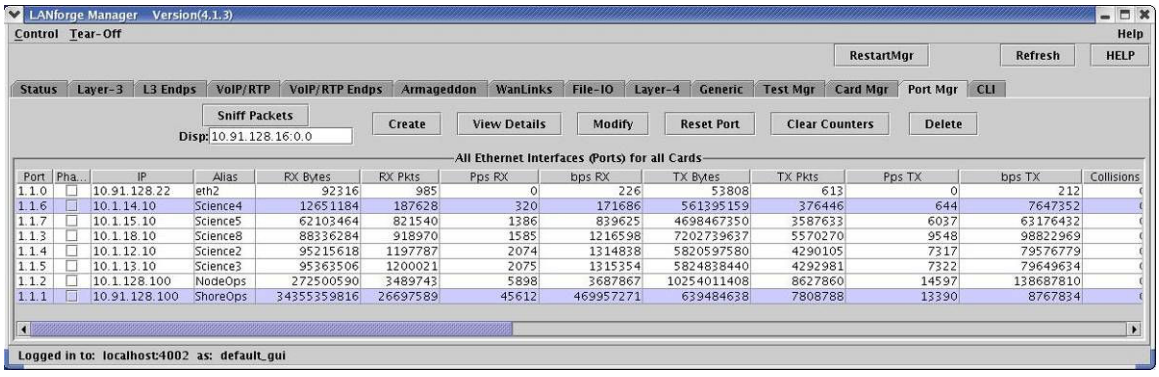
As a second test of the LanForge, a second test was performed at idle power (no external loads) to run the system (DCS and LanForge) to a stress point. The final configuration was as follows:

- Node operations: One 56kbps symmetric TCP, four 50Mbps/56kbps asymmetric TCP, two 50Mbps/56kbps asymmetric UDP. In the test setup, this connection is gigabit end-to-end, providing a lot of room for pushing data into the DCS.
- Science 2 and 3: One 56kbps symmetric TCP, two 50Mbps/56kbps asymmetric TCP, one 50Mbps/56kbps asymmetric UDP each
- Science 4: One 56kbps symmetric TCP, one 50Mbps/56kbps asymmetric TCP
- Science 5: One 56kbps symmetric TCP, one 50Mbps/56kbps asymmetric TCP, one 50Mbps/56kbps asymmetric UDP each
- Science 8: One 56kbps symmetric TCP, three 50Mbps/56kbps asymmetric TCP, two 50Mbbps/56kbps asymmetric UDP. As science 8 most accurately reflects a true science port, it was also heavily loaded.

A snapshot of the running system is shown below:



Again, all tests named “TCP’ or “test” are 56kbps symmetric. All named “hammer” are 50Mbps/56kbps TCP and all named “UDP” are 50Mbps/56kbps UDP. Again, science port 4 shows severe throughput problems. With the exception of science port 5, all of the 56kbps symmetric connections are running at nominal speeds. The majority of the 50Mbps links (both TCP and UDP) are throttled to somewhere in the 11Mbit to 37Mbit range, indicating a limit is being reached in the system. This becomes more apparent if we examine the per-interface information from the LanForge:



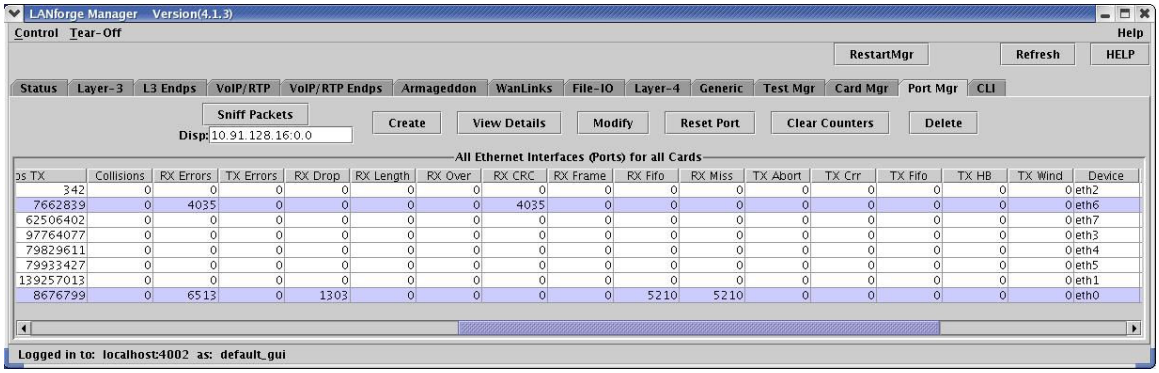
Of particular interest are the bits-per-second (bps) Rx and Tx. The shore operations interface is receiving approximately 470Mbps (!) of data and transmitting approx 8Mbps. Bear in mind this is the throughput into a single gigabit interface on the Lanforge, not the total throughput across the fiber.

The node operations interface (plugged directly into the node router) is transmitting approximately 140Mbps and receiving 3Mbps. This gigabit interface could certainly push out more data if necessary.

Port 8 seems to have reached a nominal limit of 98Mbit up and 1.2Mbit down. Although this interface on the LanForge is capable of reaching gigabit speeds, the interface auto-negotiated to 100Mbit full duplex, likely due to the cabling in the science penetrator and test pigtail. It should be noted that there is no software limitation to prevent gigabit connections at the science ports. The 100Mbit aggregate bandwidth is near or at the theoretical limit for a port operating at 100Mbit full duplex (theoretical should actually be 200Mbit less overhead…)

The other four science ports are operating on 100Mbit interfaces on the Lanforge and still pump out 60-80Mbit each from the node to the shore.

Here is additional information from the same table (scrolled right):



This shows the actual hardware level Ethernet errors reported by each interface. The second line is port 4 showing a large number of CRC errors (these counts were taken over approximately 5 minutes of full networking load), which may address the port's poor performance. The other node side ports show no errors.

The last line is for the shore operations port, which is starting to show missed interrupts and kernel buffer overruns due to the amount of traffic arriving on the port. These errors almost certainly reflect that the hardware interface on the LanForge is operating at or near full capacity.

Further information can be gathered from node Cisco routers. Here are excerpts from the data rate statistics on the four interfaces which make up the node-to-shore connection

```
mra01-01#show int Gi1/0/23
GigabitEthernet1/0/23 is up, line protocol is up (connected)
  Hardware is Gigabit Ethernet, address is 0011.218c.8746 (bia
0011.218c.8746)
(…)
  5 minute input rate 1905000 bits/sec, 2843 packets/sec
  5 minute output rate 235220000 bits/sec, 24276 packets/sec

mra01-01#show int Gi1/0/24
GigabitEthernet1/0/24 is up, line protocol is up (connected)
  Hardware is Gigabit Ethernet, address is 0011.218c.8747 (bia
0011.218c.8747)
```

```
(…)
  5 minute input rate 4041000 bits/sec, 6545 packets/sec
  5 minute output rate 70479000 bits/sec, 6676 packets/sec

mra01-01#show int Gi2/0/23
GigabitEthernet2/0/23 is up, line protocol is up (connected)
  Hardware is Gigabit Ethernet, address is 0011.218c.874a (bia
0011.218c.874a)
(…)
  5 minute input rate 3398000 bits/sec, 4877 packets/sec
  5 minute output rate 80726000 bits/sec, 7475 packets/sec


mra01-01#show int Gi2/0/24
GigabitEthernet2/0/24 is up, line protocol is up (connected)
 (…)
  5 minute input rate 0 bits/sec, 0 packets/sec
  5 minute output rate 106301000 bits/sec, 10280 packets/sec
```

The aggregate 5-minute speed from the node routers to the surface is approx 492.7Mbps. The aggregate 5-minute speed from the surface to the node is approx 9.3 Mbps. Of note, port 2/0/24 is does not seem to be receiving data. Also the link to 1/0/23 seems to receive a disproportionate amount of the data. The load balancing between the four ports is handled by Cisco's etherchannel load balancing mechanism, which is not entirely transparent.

This value likely does not represent a true maximum for the DCS. It is instead limited by the ability of the shore-side port on the LanForge to handle traffic (a consequence of dedicating a single port to the shore side).

Max throughput testing day two (20 July 2006)

In an attempt to increase the throughput across the DCS, the LanForge's ports were reconfigured as follows:

Lanforge Port	**NIC**	**Role**	**Plugged to**
Eth0	Intel Pro/1000	Topside Ops	Port 2/0/12 on topside mra91-02
Eth1	Intel Pro/1000	Node Ops	Port 1/0/11 on node router mra01-01
Eth2	Intel Pro/1000	Topside Ops2	Port 2/0/11 on topside mra91-02
Eth3	Intel Pro/1000	Science 2	Port 1/0/2 on node router mra01-01
Eth4	Digital 21143	Science 3	Port 1/0/3 on node router mra01-01
Eth5	Digital 21143	Science 5	Port 1/0/5 on node router mra01-01
Eth6	Digital 21143	Science 8	End of test pigtail
Eth7	Digital 21143	(mgmnt)	Port ??? on topside mra91-02

With the following changes:

- The management port was moved from eth2 (gigabit) to eth7 (100Mbit).
- Science 8 was moved from eth3 to eth6. The wiring length prevented gigabit speeds over this link.
- Science 2 was moded to eth3 as it is plugged directly into the node router and can achieve gigabit speeds.
- Science 4 was removed entirely as a test to diagnose the previous problems with the channel.

Rearranging the test ports necessitated essentially recreating all of the existing cross-connects. The exact configuration is not available, but similar combinations of high- and low-speed TCP and UDP connections were generated. Unlike the previous tests, all ports were left at an MTU of 1500 and the default transmit queue length (2000).

This reconfiguration allowed for two gigabit connections on the topside, and two on the node side (one on the operations subnet, and one on the science 2 subnet), theoretically allowing maximum throughput for testing.

Despite best efforts, the overall cross-connect traffic again reached a ceiling of approximately 500Mbit total traffic (aggregate traffic tx and rx from all ports). I wonder if this is a limit of our LanForge hardware (i.e. PCI bus bandwidth). Will check with Candela.