MASSACHUSETTS INSTITUTE OF TECHNOLOGY
HAYSTACK OBSERVATORY
WESTFORD, MASSACHUSETTS 01886

Telephone: 978-692-4764
Fax: 781-981-0590

December 21, 2001

TO: Distribution

FROM: Kevin A. Dudevoir

SUBJECT: Preliminary results on TCP performance between two e-VLBI workstations

1. Introduction

In this memo we present some very preliminary results on e-Vlbi data transfer performance using TCP over gigabit ethernet.

2. Testing Enviroment

The throughput measurements were performed between two potential e-Vlbi platforms whose specifications are shown in Table 1. The two machines were directly connected via a 62.5/125μm fiber cable. Subsequent tests were performed between the same two machines with one located at the Haystack correlator site and the other at the Westford antenna site. This connection is detailed in other e-Vlbi memos. These additional tests indicated comparable performance to the direct connected case and are not further documented here.
The linux 2.4 kernel used supports the extensions specified in RFC 1323 (Large Windows), RFC 2018 (Selective Acknowledgements) and path MTU discovery. The NIC used supports MTUs up to 9000 bytes. Based on this and RTT measurements between the machines, the network parameters were configured to the values shown in Table 2.
Iperf-1.2 (from the National Laboratory for Applied Network Research) was used to conduct the performance measurements. Data was collected in 5 sec intervals over a 15 minute period. No attempt at this point was made to simulate other application tasks, hence the test program had minimal contention for machine resources.

Table 1. Workstations Specifications.

Workstation 1	Workstation 2
Dell PowerEdge Server 2500	Dell PowerEdge 500SC
Pentium III 1.13 GHz	Pentium III 1.0 GHz
133 MHz front side bus	133 MHz front side bus
512 KB level 2 cache	256 KB level 2 cache
512 MB main memory	256 MB main memory
ServerWorks HE-SL chipset	ServerWorks LE 3.0 chipset
2x64-bit/66 MHz PCI slots	2x64-bit/66 MHz PCI slots
SysKonnect Model 9843 GigE NIC	SysKonnect Model 9843 GigE NIC
Linux 2.4.9-12smp (RedHat 7.1)	Linux 2.4.9-12 (RedHat 7.1)

Table 2. Network Parameters.

MTU	9000
TCP SACK	off
TCP Timestamps	off
Large Windows	on
Send buffer	286720 bytes
Recv buffer	286720 bytes

3. Performance Results

We have a raw data rate of 1000 Mb/s (the actual signaling rate on the underlying channel is 1.25 Gb/s due to the 8B/10B encoding). Every layer in the architecture imposes some overhead on this raw data rate and reduces the capacity to its higher-layer clients. The MAC layer overhead is shown in Table 3. Note that we are using a type field and not a IEEE 802.2 logical link control (LLC) field. The IP and TCP overhead is 40 bytes (no options). If we assume that the device is transferring back-to-back application data blocks of the maximum length possible, then the maximum possible throughput is given by:

$$\frac{MTU - 40}{MTU + 38} \times 1Gb/s \tag{1}$$

Table 4 shows the calculated maximum throughput values for different sized MTUs. Table 5 shows the corresponding measured throughputs. Table 6 shows the cpu loads of the two workstations during both transmit and receive for the MTU 9000 case.

Table 3. MAC Channel Overhead.

Interframe Gap	96 bits/frame (12 bytes)
Preamble/Start-of-Frame Delimiter	64 bits/frame (8 bytes)
Destination and Source Addresses	96 bits/frame (12 bytes)
Type/Length Field	16 bits/frame (2 bytes)
Frame Check Sequence (CRC)	32 bits/frame (4 bytes)

Table 4. Maximum Throughput versus MTU.

mtu (bytes)	9000	4500	1500
Throughput (Mb/s)	991.369772	982.81181	949.284785

Table 5. Measured Throughput versus MTU.

mtu (bytes)	9000	4500	1500
Mean (Mb/s)	990.061369	981.403262	704.106792
σ (Mb/s)	0.531939	0.304437	2.738205
Max (Mb/s)	991.136138	981.649372	708.205473
Min (Mb/s)	988.694203	980.636844	700.202640

Table 6. CPU Load for 9000 Byte MTU Transfer.

workstation	1	2
rx load (%)	43	60
tx load (%)	57	98