PROCEEDINGS OF SCIENCE

Investigating the e-VLBI Mark 5A end systems in order to optimise data transfer rates as part of the ESLEA Project

Matt Strong[1]
Jodrell Bank Observatory, The University of Manchester, Oxford Road, Manchester, UK
E-mail: matthew.strong@manchester.ac.uk

Richard Hughes-Jones
The University of Machester, Oxford Road, Manchester, UK
E-mail: r.hughes-jones@manchester.ac.uk

Ralph Spencer
Jodrell Bank Observatory, The University of Manchester, Oxford Road, Manchester, UK
E-mail: ralph.spencer@manchester.ac.uk

Simon Casey
Jodrell Bank Observatory, The University of Manchester, Oxford Road, Manchester, UK
E-mail: simon.casey@manchester.ac.uk

Stephen Kershaw
The University of Manchester, Oxford Road, Manchester, UK
E-mail: stephen.kershaw@manchester.ac.uk

Paul Burgess
Jodrell Bank Observatory, The University of Manchester, Oxford Road, Manchester, UK
E-mail: paul.burgess@manchester.ac.uk

Arpad Szomoru
Joint Institute for VLBI in Europe, Dwingeloo, NL
E-mail: szomoru@jive.nl

We report on the development of high bandwidth data transfers for e-VLBI at Jodrell Bank Observatory as part of the ESLEA project. ESLEA is a UK project to exploit the use of switched-lightpath optical networks for various applications, including e-VLBI, HEP, High Performance Computing and e-Health. We show how the CPU power of the Jodrell Bank e-VLBI Mark 5A end systems was limiting the data transfer rate to below 512 Mb/s. Both of the Jodrell Bank Mark 5A end systems have now been upgraded and can now transfer e-VLBI data to JIVE at the required data rate of 512 Mb/s.

Lighting the Blue Touchpaper for UK e-Science - Closing Conference of ESLEA Project
The George Hotel, Edinburgh, UK
26-28 March, 2007

[1] Speaker

1. Introduction to VLBI, and e-VLBI

Very Long Baseline Interferometry (VLBI) is a technique for creating high resolution radio maps using radio telescopes located around the world (and even in space). A defining feature of VLBI is that, historically, due to the large distances involved, the telescopes store data on magnetic tape, or more recently computer disks as they cannot be connected directly. These tapes or disks are then shipped to the correlator, played back, correlated and Fourier transformed in order to create the high resolution images. Recently however, there has been a drive to upgrade the VLBI system to a real time instrument, e-VLBI (e-VLBI in Europe is being developed with funding of the EU project EXPReS). The use of computer networks and the internet are ways of connecting radio telescopes around the world together, and so VLBI astronomy can be performed in real time.

The current VLBI system employs the Mark 5A disk-based recorder [1], which records the astronomical data collected at the telescope to large disk packs with capacities of several hundred gigabytes each. The core of the Mark 5A is a 1.2 GHz standard PC running Linux. The PC contains two interface boards; a StreamStor card[*] for high speed disk reading and writing and an I/O board [2]. As the Mark 5A is simply a custom designed PC which interfaces to the VLBI formatters and disk packs, it is possible to retro-fit a gigabit Ethernet card via the PCI bus. The Mark 5A control software is capable of reading/writing data via the Ethernet card, in a similar way to how it communicates with the disk packs and formatter. It is therefore possible to establish a direct ‘link’ between the telescope and the correlator and perform real time e-VLBI.

Current Production Goals of e-VLBI

The Mark 5A units are capable of recording data at rates of up to 1 Gb/s, and whilst the Ethernet interface can run at 1 Gb/s, it is not possible to achieve transmission of telescope data at 1 Gb/s. This is due to the fact that the data has to be encapsulated within TCP or UDP packets, which then have to be encapsulated in IP packets and finally in Ethernet packets. Each level of encapsulation adds a little more data that needs to be transmitted, and so for 1 Gb/s of telescope data, there would be 10% more data created by the encapsulation and hence would not be transmitted through a 1 Gb/s Ethernet card. Owing to the nature of the VLBI formatters, and current technical constraints, data rates have to be a power of 2 (32, 64, 128, 256, 512, 1024 Mb/s) and so the next speed down is 512 MBits/s which is technically achievable over a 1 Gb/s link. It is thus the current goal of the e-VLBI community to reliably run e-VLBI experiments at 512 Mb/s, and then to develop the technology and networks in order that this speed can be increased to gigabit levels.

The e-VLBI Network

The Mark 5A units, and other e-VLBI PCs stationed at the telescopes and correlator communicate with each other over the European wide production network for academic research and development, known as GÉANT 2 (with the exception of the Westerbork telescope which has its own fibre connection to the correlator). The Mark 5A end systems are connected to the GÉANT 2 production network through their local and national research and education networks (NRENs). In addition to the GÉANT 2 production network, Jodrell Bank Observatory in the UK also has two dedicated optical links between Jodrell Bank and JIVE. These links are routed via UKLight, and its peering ability with SURFnet and NetherLight. There are dedicated

[*] Made By Conduant

1 Gb/s and 630 Mb/s optical connections between the telescopes at Jodrell Bank and the correlator at JIVE.

2. The 500 Mb/s Bottleneck

Currently, EVN e-VLBI operate a TCP based system, and optimisation of this system is necessary if e-VLBI is to work at the highest data rates (512 Mb/s is the current goal). In e-VLBI system testing, two of the five participating stations (Onsala an Westerbork) have been able to achieve 512 Mb/s using their current Mark 5A machines, but 512 Mb/s transfer between Jodrell Bank and the correlator (located at JIVE in the Netherlands) could not be achieved despite identical hardware and spare capacity on the network links. e-VLBI data transfers of 500 Mb/s could be achieved on both the production link and the 1 Gb/s dedicated UKLight link provided by the ESLEA project, but this point proved to be a bottleneck with the existing hardware. This bottleneck was thought to be caused by the Jodrell Bank Mark 5A system and as such investigation of its performance was necessary. The results of this study are detailed in the following sections.

3. Mark 5A UDP Transfers over UKLight

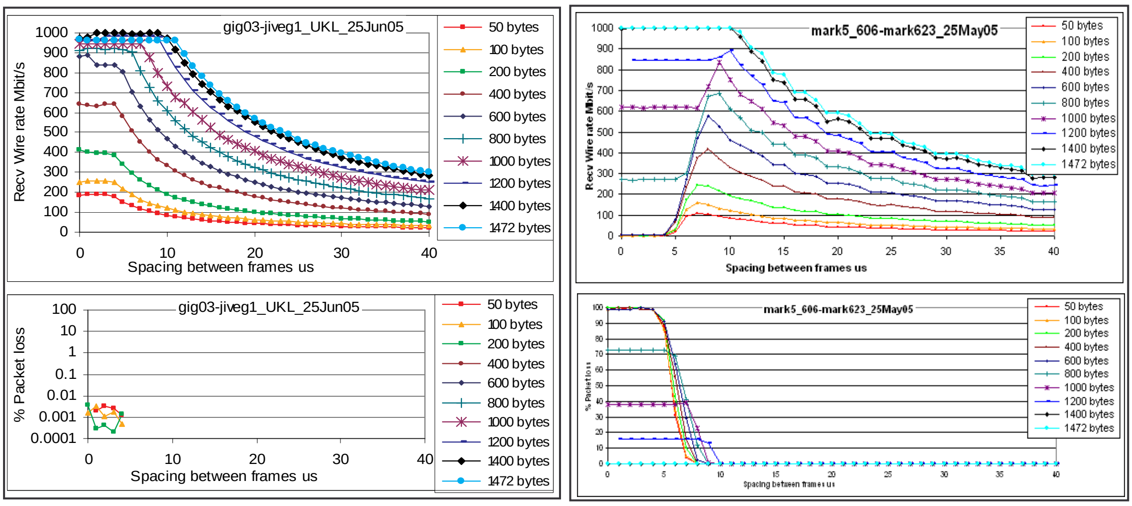

Figure 1 - UDP_mon transmission tests. Left Panel: UDP throughput and packet loss between Jodrell Bank and JIVE using high performance network machines. Right panel: UDP throughput and packet loss between Jodrell Bank and JIVE using the Mark 5A machines.

The performance of the Mark 5A end systems was first investigated by transferring UDP data (using the UDP_mon software package†) between Jodrell Bank and JIVE and comparing these results with those from a high performance network machine. Figure 1 shows the results from this study. The left panel shows the results from the high power network machine, and we can see good throughput for all inter-packet spacings, whilst only a very small amount of packet loss is detected at small packet sizes and inter-packet spacings. The right hand panel of Figure 1 shows the corresponding graphs for the Mark 5A machines. It can be seen that the UDP throughput shows its normal signature for larger inter-packet spacings. However, the throughput falls off dramatically for low inter-packet spacings for all packet sizes. Also, from the lower graph it is evident that there is dramatic packet loss corresponding to this loss in throughput. A

† Developed by Richard Hughes-Jones

common cause of such a loss of throughput can be the CPU speed of the machines. Below we investigate the CPU performance of the Mark 5A machines to ascertain if it is having an adverse effect on the e-VLBI data transmission.

4. Analysis of the CPU Performance of the Mark 5A End Systems

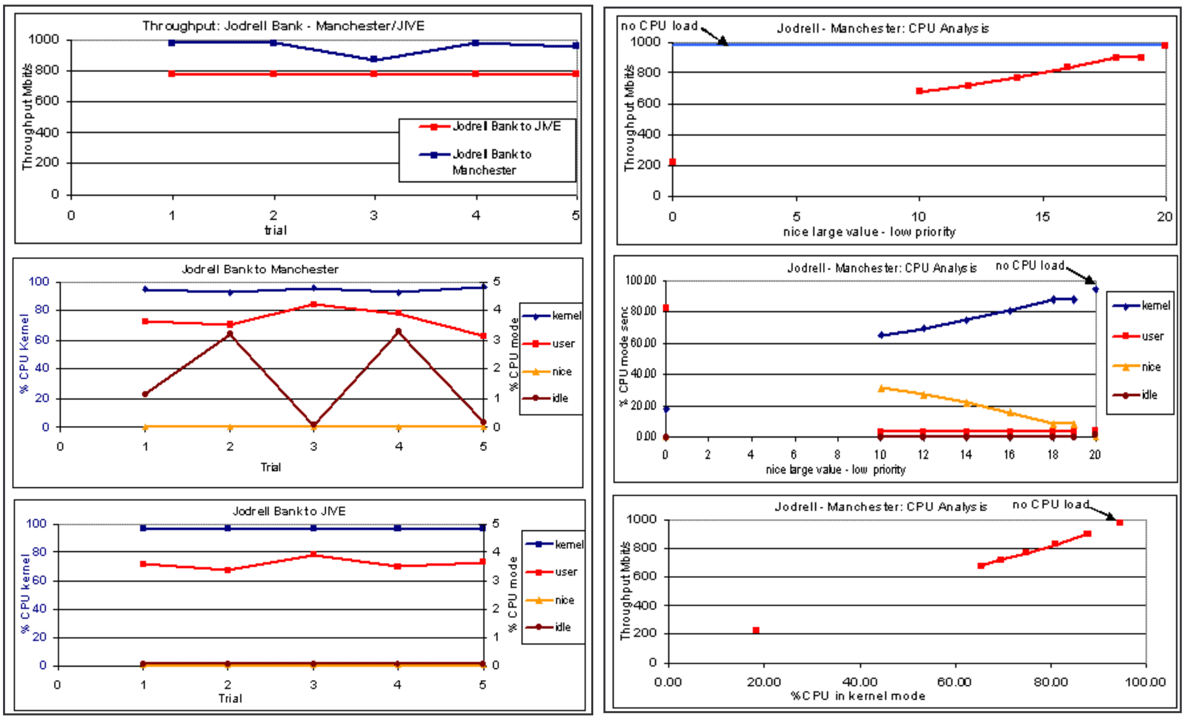

Figure 2 - CPU analysis of the Jodrell Bank Mark 5A. Left panel: Network and CPU performance whilst performing an iPerf TCP transmission between Jodrell Bank and Manchester, and Jodrell Bank and JIVE. Right panel: Network and CPU performance whilst performing an iPerf TCP transmission between Jodrell Bank and Manchester and also running a CPU intensive task.

In order ascertain if the Mark 5A CPU speed was having an adverse effect on e-VLBI transmission, its usage was examined prior to and after each test. In addition to CPU load, network interface, IP, UDP (via UDPmon) and TCP (via iPerf) statistics were also measured just prior to, and just after each test. Thus, this allowed much of the resources to be measured. Tests were performed between Jodrell Bank and Manchester, and then Jodrell Bank and JIVE, both on the dedicated optical network.

For a single, memory to memory TCP stream, the connection between Jodrell Bank and Manchester showed a transmission rate of 950 Mb/s (Figure 2, top left) and a CPU usage of 94.7% kernel, 1.5% idle (Figure 2, middle left), whilst the connection between Jodrell Bank and JIVE showed a transmission rate of 777 Mb/s (Figure 2, top left) and a CPU usage of 96.3% kernel, 0.06% idle (Figure 2, bottom left). Thus, it appears that when transmitting between Jodrell Bank and Manchester, the Jodrell Bank Mark 5A machine just has enough CPU to send the data at line rate. However, when transmitting between Jodrell Bank and JIVE, the Jodrell Bank Mark 5A does not appear to have sufficient CPU power to drive the network at line rate. e-VLBI transfers are obviously not memory to memory transfers, as the data has to be processed in the Mark 5A machine, resulting in it passing through the PCI bus a number of times. As such, addition CPU is necessary to perform such processing.

Due to the fact that we cannot simulate exact e-VLBI transmission as the receiving machine is not a Mark 5A, we simulated the effect of the Mark 5A processing by adding a CPU intensive task to the sending Mark 5A machine (at Jodrell Bank). Thus, we could then investigate how the throughput and CPU usage varies with respect to this additional task.

From Figure 2 (top right) it can be seen that the throughput of memory to memory TCP flows between Jodrell Bank and Manchester fell from 950 Mb/s with no CPU load process to 900 Mb/s when the CPU load process had the lowest "nice" priority of 19 and to 675 Mb/s when the "nice" priority was 10. The "nice" priority range runs from -20 at its highest to +19 at its lowest, with normal user priority 0. In addition to this, the middle right graph in Figure 2 shows that the available CPU in Kernel mode falls rapidly to approximately 60% when the "nice" priority increases from 19 to 10. The bottom right graph in Figure 2 shows how the throughput of the TCP stream is related to the available CPU power and shows the throughput falling as the available CPU decreases. These graphs clearly show that the addition of a CPU intensive task has an adverse effect on the TCP transmission speed. It is clear that the Jodrell Bank Mark 5A machine does not possess enough CPU power to drive the network at adequate speeds whilst performing other processing. For this reason, the Jodrell Bank Mark 5A was upgraded with an Asus NCCH-DL motherboard, Intel Xeon 2.8 GHz processor and 1 GB of server specification SDRAM. The tests above were then repeated with the new upgraded Mark 5A machine and the results are given in section 5.

5. CPU Performance Tests of the Upgraded Mark 5As

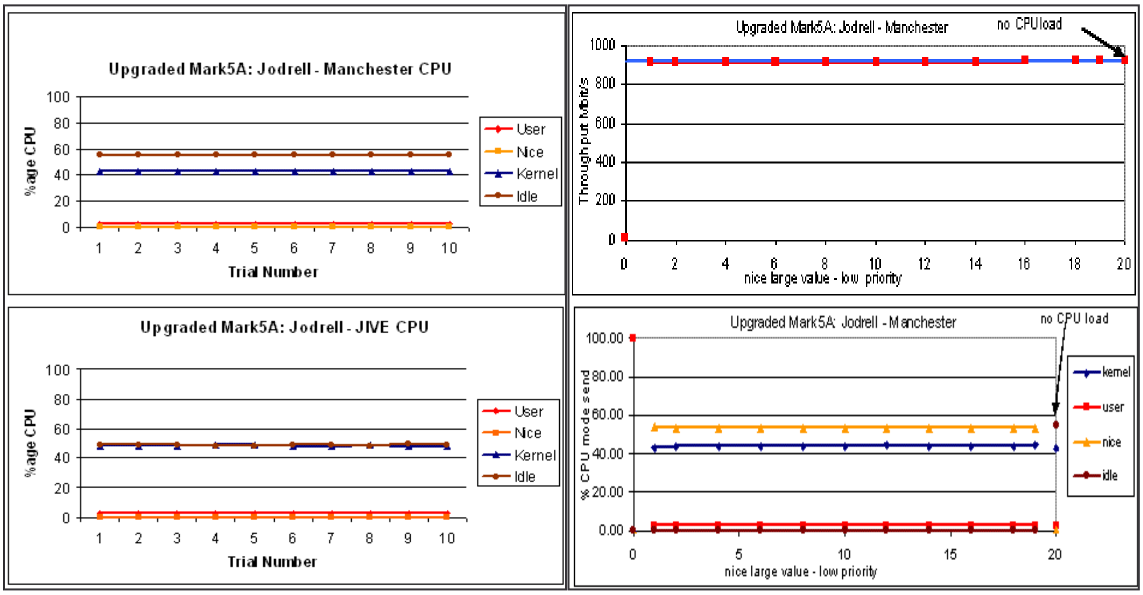

Figure 3 – CPU analysis of the upgraded Jodrell Bank Mark 5A. Left panel: CPU performance whilst performing an iPerf TCP transmission between Jodrell Bank and Manchester, and Jodrell Bank and JIVE. Right panel: Network and CPU performance whilst performing an iPerf TCP transmission between Jodrell Bank and Manchester and also running a CPU intensive task.

Figure 3 shows the CPU performance of the upgraded Mark 5A machine. It is easy to see that the upgraded Jodrell Bank Mark 5A performs extremely well in these tests. Indeed, it maintains a near line rate memory to memory TCP stream whilst the "nice" priority of the CPU intensive is varied over its full range. Indeed, the bottom right graph in Figure 3 shows the upgraded Mark 5A machine to be using less than 50% of its CPU in transferring the data and running the CPU intensive task.

After these tests were performed, standard e-VLBI tests were performed with this upgraded Mark 5A machine, and it achieved e-VLBI data transmission at a rate of 512 Mb/s over the dedicated optical link immediately. It seems obvious that this Mark 5A upgrade has had a positive effect on its data transmission, and now it can achieve 512 Mb/s e-VLBI data transfer as required.

6. Conclusions

From the above tests it is clear that the Jodrell Bank Mark 5A machine did not have sufficient CPU power to transfer e-VLBI data at 512 Mb/s. This result is surprising as both the Onsala and Westerbork Mark 5A machines have been able to transfer eVLBI data at 512 Mb/s. So what is the difference between these machines? Before the Jodrell Bank Mark 5A was upgraded, the specification and components of all the Mark 5A machines were the same. They were Intel P3, 1.2 GHz machines with 256 MBytes of RAM.

It is noted that the difference between the transmission speeds between the Onsala and Westerbork Mark 5As and JIVE, and the initial Jodrell Bank Mark 5A and JIVE was at the 10% level. The specified CPU speeds of the Mark 5A machines are only accurate to ~10% level, and as such, this inaccuracy could be responsible for such a bottleneck. Indeed, if the initial Jodrell Bank Mark 5A's clock speed was 10% lower than its specification, and the Westerbork and Onsala Mark 5A's were 10% higher, this leaves a shortfall on CPU power of ~240 MHz. Such a shortfall in CPU power could have easily resulted in the 512 Mb/s threshold being unattainable from the Jodrell Bank Mark 5A.

Regardless, it is certain the upgrade to the Jodrell Bank Mark 5A has had a dramatic effect on its performance. Indeed, with the upgraded Mark 5A, Jodrell Bank obtained reliable 512 Mb/s data transmission to JIVE immediately. One of the main problems with the Mark 5A machines is that they require CPU power to transfer the telescope data across the PCI bus twice. This results in the Mark 5A machines needing more CPU to drive the network at the necessary rates. Indeed, ultimately the goal of e-VLBI is to transfer data from the telescopes to the correlator at over 1 Gb/s, and as such it is unclear as to whether the Mark 5A machines can manage these rates.

References

[1] Alan Whitney, *Mark 5A disk-based gbps vlbi data system*, viewed 19th July 2006, http://web.haystack.mit.edu/mark5/paper.pdf

[2] Dan L. Smythe. *Mark 5A Memo #007.1. Mark 5A memo series*, viewed 19th July 2006, ftp://web.haystack.edu/pub/mark5/index.html