High Rate Internet Data Transfer for eVLBI

R. Spencer, R. Hughes-Jones and S. Casey

The University of Manchester

Abstract

Very Long Baseline interferometry (VLBI) is a well established technique in radio astronomy for the study of celestial sources at high angular resolution. VLBI is gradually moving to the point where Gbps data rates are becoming routine. A number of experiments have shown that the internet can be used at data rates of several hundred Mbps on production networks. However the use of switched light paths should enable higher data rates to be achieved. VLBI is resilient to packet loss, but not to loss of bandwidth. The statistics of packet loss as found by recent tests and the effects of loss on VLBI performance will be discussed. Recent measurements comparing the UKLight connection to JIVE in the Netherlands will be compared with tests on the production network.

1. Introduction

Very Long Baseline Interferometry (VLBI) is used by radio astronomers to obtain detailed images of cosmic radio sources. The technique was first developed in the late 1960's in pioneering experiments which involved flying atomic clocks across the Atlantic and using limited bandwidth (< 1 Mbps) computer data tape to record the data [1]. Rapid developments took place as the video recording industry grew (VHS cassettes were favoured in the 1980's) with increases in data rates to that of the MkIV video tape system capable of routine operations at 256 Mbps in the late 1990's and then to the MkV 1 Gbps disk recording system in use today [2].

The background to VLBI and eVLBI will be described first, followed by discussion of experiments on the production network and the effects of packet loss on VLBI data. Finally recent results on a comparison between UKLight and production connections from Manchester to Dwingeloo will be discussed.

2 How it works

2.1 Interferometry

The technique relies on the fact that the cross-correlation of the electric field measured on a plane surface is the Fourier transform of the sky brightness. This is a theorem in optics (the van Cittert-Zernicke theorem, discovered in the 1930's [3]) but it was not until the invention of the aperture synthesis technique in the 1960's by Martin Ryle and Tony Hewish [4] that it was applied to produce images of celestial radio sources by merely Fourier inverting the cross-correlation data [1]. In other words an interferometer produces a measurement of one Fourier component of the radio source's angular structure. By bringing together many Fourier components a representation of the structure of the source can be found.

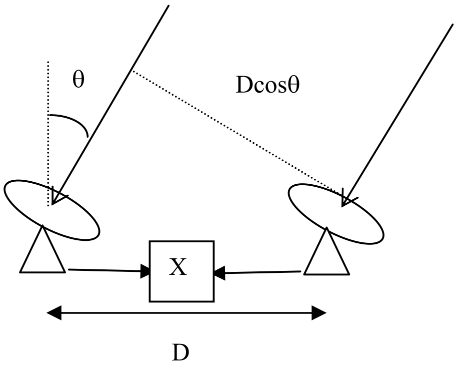

Figure 1 *Schematic of a simple interferometer. Signals from direction θ are brought together at a central site and multiplied, producing an output proportional to the cross-correlation of the electric field arriving at each telescope.*

Figure 1 shows a simple interferometer made of 2 telescopes separated by the baseline D. Receivers at the focus of each telescope convert the radio waves to electrical signals which are brought together to be multiplied and averaged This can be done by analogue electronics or digitally. The multiplier forms one 'delay' channel of a correlator, where the delay between the signals is incremented for each channel of the correlator. The output has amplitude and phase and gives one Fourier component, varying D gives further components. These components are also called visibilities. The angular resolution increases with D, and the highest resolution is obtained by using global baselines.

As the Earth rotates the angle θ will change and hence the projected baseline Dcosθ changes with the angle of the source. The projected baseline defines an 'aperture plane' in two dimensions and it can be shown that an ellipse is produced in the aperture plane (also labelled as the u,v plane) as the Earth rotates. Each interferometer pair forms a separate ellipse, and a combination of N telescopes produces N(N-1)/2 ellipses. Figure 2 shows a set of ellipses produced by 8 telescopes of the European VLBI Network (EVN) over a 14-hr observing period.

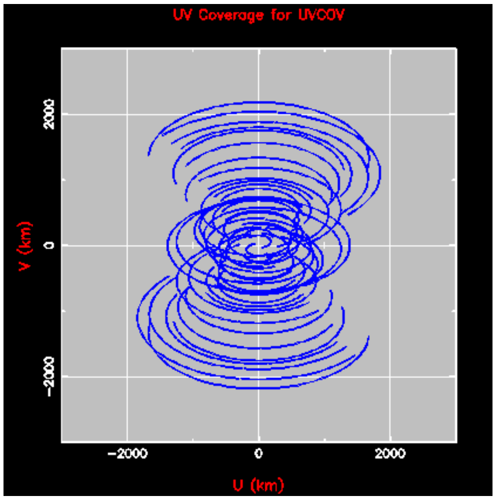

Figure 2 *Aperture plane coverage produced by EVN for a source at declination 32 degrees. The cut off in the ellipses occurs because the source is below the horizon at that time, the reflection symmetry through the origin occurs because the sky brightness is real hence the visibilities are Hermitian. The array effectively synthesises a telescope 2000 km in diameter. The fuller the uv plane coverage the better the imaging quality.*

2.2 VLBI

The highest resolution requires the longest baselines. Telescopes can be separated by 1000's km in VLBI and data transfer to the correlator is traditionally done by recordings on magnetic tape. The UK participates in European VLBI experiments as a member of the European VLBI Network (EVN), with data being correlated using the EVN data processor at the Joint Institute for VLBI in Europe (JIVE), Dwingeloo, The Netherlands [15]. Other correlators exist including a similar correlator to that at JIVE at MIT, Haystack, in Massachusetts, USA. In the last couple of years EVN has moved over to disk recording (the MkV system), with a major improvement in reliability.

The correlation process occurs off-line and can be months after the observations. Faults can therefore not be corrected in time and so reliability has been an issue. A dedicated array of telescopes specifically for VLBI - the VLBA - was built in the USA in the 1990's and through continuous operation (not possible for European telescopes) and good engineering practice the array achieved good reliability. In the last few years the use of ftp file transfer of short sections of data from the European telescopes to the JIVE correlator has improved reliability. However substantial improvements to performance and usability could accrue if true real time operation can be accomplished. The use of high data rate transfers on research network infrastructures in eVLBI offers a way of achieving true real time connection over transcontinental baselines.

3 eVLBI

As well as improving reliability, the use of the internet for real time data transfer offers a number of advantages for VLBI science:

- The bandwidths possible, and hence the sensitivity that can be obtained, will be eventually significantly higher than that of disk-based system. The sensitivity in radio astronomy is proportional to the square root of the bandwidth for a given observing time for continuum sources. Receivers are approaching the quantum limit in noise performance and larger telescopes are expensive so the most feasible way of improving sensitivity is to increase the bandwidth.

.

- Real-time VLBI will be vastly more reliable than the current system in which data are often not correlated until several weeks after recording
- Permanent connections via fibre and real-time VLBI will result in a major culture change:

- Regular and frequent observations of variable radio sources
- Rapid response to targets of opportunity: supernovae, Gamma ray burst sources, microquasar transients etc.
- Move away from only three sessions a year to regular VLBI observations
- Dynamic, central scheduling of telescopes and correlators
- Simultaneous continuum and multi-line spectroscopy.

3.1 eVLBI Tests and Milestones

The advantages have been taken on board by the VLBI community and a development programme has been instigated by the directors of EVN observatories. A number of tests and demonstrations have taken place over the last few years:

- September 2002: Jodrell Bank (JBO)-Westerbork fringes were obtained for the iGRID 2002 exhibition in Amsterdam. Peak transfer rates of 500 Mbps were obtained to Amsterdam [5]
- October 2002 – July 2003: various small-scale tests were undertaken and ftp-vlbi implemented at data rates of a few 10's of Mbps
- July 2003: Westerbork telescope connected at 2.5 Gbps to Dwingeloo
- October – December 2003: Data transmission tests between Manchester and Dwingeloo achieved more than 900 Mbps – see this paper
- November 2003: An international baseline, Onsala (SE) – Haystack (USA) was used, producing eVLBI fringes only 15 minutes after observations were made.
- November 2003: Onsala Space Observatory connected at 1 Gbps.
- January 2004: First (disk-buffered) eVLBI image using data from Jodrell Bank, Onsala and Westerbrok. Data were ecorded on MkV and transmitted over the normal internet connection from JBO (using 155 Mbps connection from JBO → Manchester) and over dedicated links from Onsala and Westerbork. Data were then received on MkV systems at JIVE, buffered on disk and played back into the correlator.
- April 2004: Real-time fringes Onsala-Westerbork (no disks) – with data streaming directly from the telescopes to the correlator.
- 28 April 2004: First image from a real-time eVLBI session involving JBO, Onsala and Westerbork**.** No disks were used
- September 2004: Four telescope real-time eVLBI, Fringes to Torun and Arecibo were found, first eVLBI science session, producing an image of a OH maser source.
- December 2004: connection of JBO to Manchester by 2 x 1 GE links. eVLBI tests run between Poland, Sweden, UK and Netherlands at 256 Mbps., but with variable success, only runs at 128 Mbps were reliable
- January 2005: First "dedicated light-path" eVLBI between Australia and JIVE. Data rates of 450 Mbps were achieved from observations of the Huygens descent to Titan, using non-VLBI equipment.
- February 2005 TCP and UDP memory – memory tests at rates up to 450 Mbps (TCP) and 650 Mbps (UDP). Tests showed inconsistencies between Red Hat kernels, rates of 128 Mbps only obtained on 10 Feb, though Haystack (US) – Onsala (Sweden) ran at 256 Mbps
- 26 May 2005. Successful tests with the Cambridge, Jodrell, Onsala, Torun, Westerbork and Arecibo telescopes connected to the JIVE correlator. Successful and reliable real time fringes were obtained at 64 Mpbs and also at 128 Mbps without Arecibo.

The above tests have revealed that even though there are limitations in the MkV VLBI equipment used there are further restrictions on the data rates currently achievable on the production networks. Nevertheless the data rates are sufficient for good science to be obtained, and eVLBI has been advertised as available for astronomers to use in the latest EVN 'Call for Proposals' [6].

4. Network Tests

It is still unclear how VLBI should make the best use of the internet: data rates may be limited by the end hosts in hardware and software, by the local area network, by the access links between national and international networks and not least by the protocols used. We therefore initiated tests on the network, independent of the VLBI tests above, since we would then not depend for example on the availability of telescopes An important parameter is the effect of packet loss on the data, and this problem is addressed in this paper. Our data rates are high compared with the average Internet user and so we must also be aware of the possibility of denial of service to others.

4.1 Packet Loss Measurements

An investigation of the link from the University of Manchester to JIVE in Dwingeloo was undertaken as an undergraduate project in the autumn of 2003. The link used the SuperJANET4 academic network in the UK to connect to London, then GEANT to Amsterdam, followed by SURFnet to JIVE in Dwingeloo.

There are two main protocols in common use on the internet: Transmission Control Protocol (TCP) which ensures bit-wise correct data movement and is used by most ftp systems, and User Datagram Protocol (UDP) which has little error handling, does not guarantee delivery and requires the user application to deal directly with any end-to-end communication problems. With both transport protocols data are divided into sections and sent as packets. The larger the packets the higher throughput of user data, as the protocol overhead is a smaller percentage. The maximum size of the packet (MTU) is limited by the length that can be accommodated by routers in the link. In our case this was 1500 bytes, allowing 1472 bytes of user data. Data were placed on the LAN using 1-Gigabit Ethernet connections. Data rates of close to 1 Gbps can be achieved provided appropriate network interface cards (NIC) and machines are used [7].

TCP/IP produces a bit wise correct transfer and tries to be fair for other users. Missing packets are interpreted as congestion and the rate is halved. This can result in a highly variable transmit rate, but with no missing or corrupted data and perhaps explains the various data rates obtained in some of the eVLBI experiments above. UDP however will transmit at a rate determined by the user and the available bandwidth, and has no acknowledgement. Packets can therefore be lost or out of order with no effect on the transmit rate.

The network tool UDPmon [8] was used to test the link. It sends packets at carefully spaced intervals and measures data rates and loss at the receiver as a function of packet size and inter-packet interval. Figure 3 shows the data rates achieved using a dual Xeon 2 GHz machine at Manchester and the 1.2 GHz PIII MkV machine at Dwingeloo in tests made on 11th November 2003 (Man-Dwing) and 13th November 2003 (Dwing-Man). Near wire rates were achieved at maximum using 10^6 packets per measurement in UDPmon, simulating near-continuous data transfer. The fall off as 1/(packet spacing) occurs because the link is waiting for data, the flat part of the curve indicates that the link or computers are limiting the rate. Figure 4 shows that it is in this flat region that packet loss occurs.

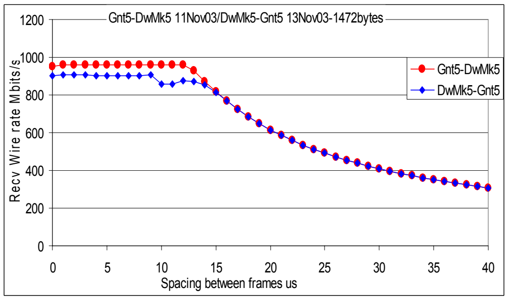

Figure 3 *Received data rate as a function of packet spacing on the Manchester-Dwingeloo link, and in the reverse direction. The two curves show that the link is asymmetric, and reflects the different compute power available at each end.*

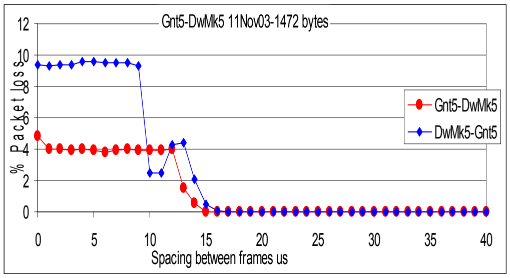

Figure 4 *Packet loss versus packet spacing on the Manchester-Dwingeloo link*

Packet loss can occur due to insufficient processor power or bus capacity in the end machines, or by congestion in routers on the link. Figure 5 shows the traffic on the 1 GE link from the university to the Net North West router. The rates are averaged over 5 minutes. Our tests, with average data rates of ~400 Mbps, are clearly visible and so congestion could occur when our peak data rate reaches ~900 Mpbs. Note that this is much higher than the rates achieved in the eVLBI data tests.

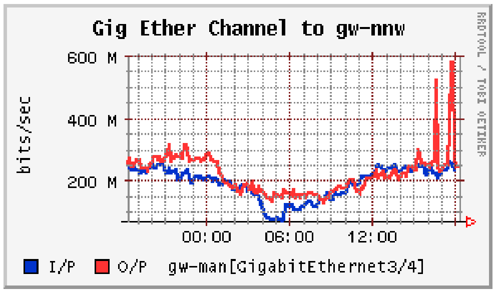

Figure 5 *Traffic to the Net North West router in Manchester showing the effect of the tests at around 17:00. The traffic rates are averaged over 5 min.*

4.2 Effect of Packet Loss on VLBI data

Loss of data will cause a decrease in signal to noise (S/N) in VLBI observations. In normal circumstances S/N will be proportional to $\sqrt{(1-f)}$ where f is the fraction of packets lost. However if the loss of data is sufficient for the correlator to lose synchronisation rather more can be lost. The JIVE correlator checks parity of each 9-bit (8 bit plus a parity bit) VLBI byte. If more than 10 % of the bytes per VLBI frame are wrong then the whole frame of 2500 9-bit bytes is rejected. Luckily the correlator can flywheel synchronisation over to the next VLBI frame, but if successive frames are rejected then synchronisation is lost. This gives VLBI systems some resilience to data loss - a consequence of having to deal with drop-outs on tape systems. In fact calculations estimate that packet loss as high as 2% can be tolerated without serious degradation of VLBI data [9].

This calculation is only true if packet loss obeys a Poisson distributed statistical process, i.e. if successive packet loss is independent and obeys normal counting statistics. However congestion is likely to result in correlated bursts of packet loss resulting in a long term correlation, whereas the distribution of intervals for a Poisson process follows a falling exponential [10].

Figure 6 shows the cumulative distribution of intervals between lost packets found in tests on 4th December 2003 when a total of 1409 packets were lost in a 0.6 sec run. The mean time between lost packets was 424 μsec, in good agreement with 394 μsec found from the fitted exponential. There does seem to be an excess of events for bins greater than ~100 (1200 μsec) showing some evidence of long-term effects, however we can conclude that packet loss obeys Poisson statistics to a reasonable approximation for times up to around 1 msec.

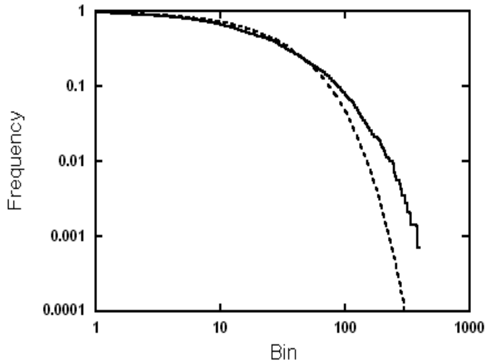

Figure 6 *Normalised cumulative distribution of packet loss (solid curve) and a falling exponential fitted to the data (dashed) versus bin number. Each bin is 12 μsec wide.*

Long term effects are expected in such data. Analyses of world-wide-web traffic have shown that the flow is self-similar [11,12]. Traffic occurs in bursts, and file transmission times have more events with long transmission times than expected often obeying a power law. Packet loss, if related to congestion, is expected to show similar behaviour, though we have little evidence of a power law tail in the distribution of packet loss in figure 6. However these data were from one run only, more runs covering a wider variety of link conditions might show self-similar effects more clearly.

It should be noted that the high packet loss shown in figure 4 was a relatively rare event. Several runs were made with negligible packet loss, which would not affect VLBI, but nevertheless cause a back off in TCP rates.

5 TCP or UDP?

The question of which protocol TCP or UDP, to use for eVLBI is important. TCP gives reliable data transfer and is fair to other users, UDP can achieve high throughput, but could in some circumstances lead to denial of service for other users.

In normal operartion standard TCP/IP assumes that any packet loss indicates congestion on the netwook and so dramatically reduces the sending rate. This 'collision avoidance' behaviour has a much more serious effect on eVLBI sensitivity than packet loss [9] in a UDP stream. As an example of this tests between MkV units at JBO and Dwingeloo using IPerf to run TCP/IP on the production link (1500 byte packets) typically achieve rates of 300-400 Mbps. These rates can be compared with tests on UKLight from Manchester to Dwingeloo, achieving rates of around 600 Mbps, rising to 864 Mbps with Jumbo (9000 byte) packets.

Packet loss, whether due to limits on the receiving machine or the network, reduces the bandwidth in TCP. The recovery time on standard TCP can be several minutes to hours depending on the length of the link, resulting in long period of low bandwidth. Variants of TCP such as High-speed TCP, Scalable TCP and FAST all address this to a reasonable extent. However TCP's completely reliable service model and the retransmission mechanism used to implement it may be inappropriate, in which case DCCP or TSUNAMI may be more useful for eVLBI. The MIT Haystack group has developed a variant of RTP over UDP (VSI-E) which may be best suited to eVLBI using the MkV recording system [13].

6 Tests on Switched Light Paths

An answer to the congestion problem on production networks may be found with the advent of switched light path systems. The sources and destination (telescopes and correlator) have fixed locations and therefore ideally suited to the concept of switched light paths. Recently the ESLEA (Exploitation of Switched Lightpaths for eScience Applications) [14] project has been initiated in the UK which makes use of the UKLight dedicated research network. UKLight connects Manchester though Warrington to London and then on Amsterdam using dedicated 1 Gigabit SDH circuits. Dedicated 1 Gigabit Ethernet links from SURFnet form the onward connection from Amsterdam to JIVE.

We were able to make UDPmon tests on this network, and also able by using different hosts to make similar tests using the production network (Super JANET and GEANT).

Figure 7 shows the results of tests between Gig03 (a 2.0 GHz hyper-threaded dual Xeon processor with a 133 MHz 64 bit PCI-X bus) in Manchester and a similar machine, JIVEgig1, in Dwingeloo. The machines were connected via UKLight. The plots show the results for a number of different packet lengths, the curves for packet sizes of 1472 bytes are the most relevant for VLBI. The throughput limitation for packets smaller than 600 bytes is due to the time to move the data over the PCI bus in the end host. Packet loss (at the 0.003 % level) was only observed for packets of 200 bytes and less with packet spacing closer then 4 µs. Note that there was no packet re-ordering on this link.

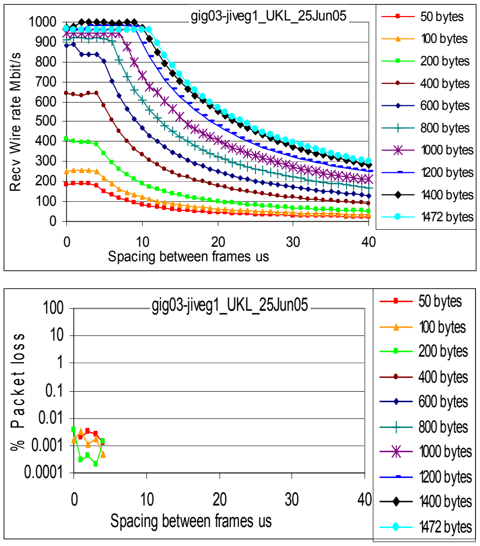

***Figure** 7 Tests on the UKLight switched light-path connection between Manchester and Dwingeloo. a) (upper) shows throughput and b) packet loss as a function of inter-packet spacing. Maximum size packets can reach full line rates with no loss, and there was no re-ordering (plot not shown).*

Figure 8 shows a similar plot for Gig6, (a 2 GHz hyper-threaded dual Xeon machine) also in Manchester but connected to JIVEgig1 via the production network.

At first sight the production link seems to be performing as well as UKLight but packet loss was slightly higher on the production route. However packet reordering on the production link (Figure 9) was significant and this would cause an extra processing load for e-VLBI transfers. Repeated tests on the UKLight link have shown no reordering.

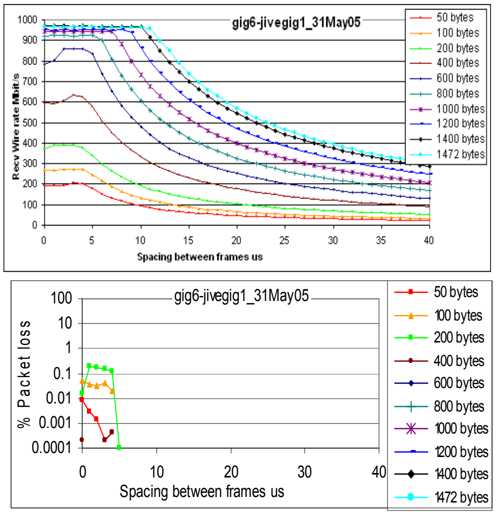

***Figure 8** Tests on the production network between Manchester and Dwingeloo. As before the upper graph shows throughput, the lower packet loss. Small (0.2%) packet loss was seen, and re-ordering of packets was significant (see figure 9)*

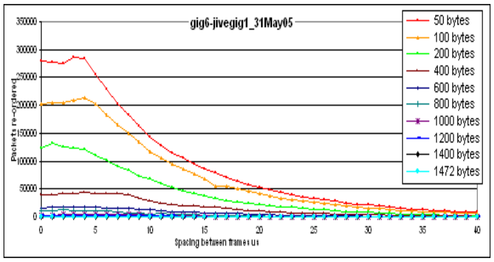

***Figure 9** Reordering on the production link.*

Figure 10 shows the tests between the MkV VLBI recording systems at Jodrell Bank Observatory and JIVE via UKLight.

These are P3 machines running at 1266 MHz, at considerably lower performance than the Gig series of test machines as can be seen by the high loss rates for small and medium sized packets. The receiving CPU was running at around 80% capacity for inter-packet spacings of less than 12 µsec.

Nevertheless peak rates of 999 Mbps were seen, with low packet loss for 1472 byte packets. However it is interesting that no re-ordering was seen, consistent with the other experiments on UKLight.

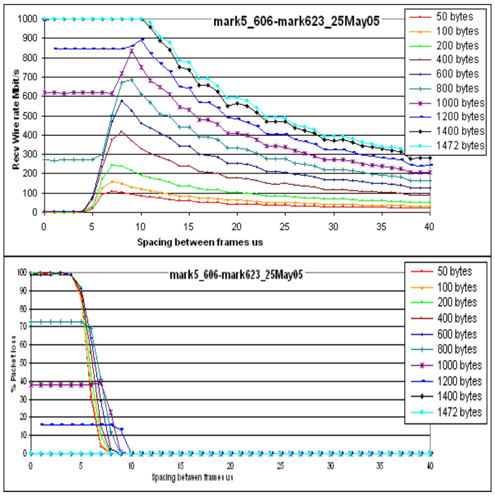

***Figure 10** Tests on the UKLIght connection using MkV VLBI terminals.*

7 Discussion and Conclusions

It is clear that the data rates achievable on production links are much higher with UDPmon and IPerf than has been possible to obtain in eVLBI experiments (typically only 128 Mbps reliably). The main difference is that VLBI data are continuous and use TCP/IP. Congestion could cause TCP to reduce bandwidth over a significant fraction of the observing period due to the long recovery time of long links in TCP. In order to investigate packet loss further, new software to generate continuous data flows with UDP and incorporating monitoring routines is being developed. However at the moment it would appear that VLBI data transfer using UDP is capable of giving higher bandwidth and hence better sensitivity than TCP. Further work on TCP variants is needed before we can arrive at the most effective data transport protocol for eVLBI. This is a research topic in the ESLEA project.

Our experiments show that the assumption of Poisson statistics for the distribution of packet loss is a good approximation, but further work is needed to clarify expected power law behaviour at long intervals. Furthermore packet loss definitely attributable to the network is infrequent, occurs in bursts, and has therefore not been subjected to the same analysis. At some point packet loss will cause correlator faults and it is important that this level be defined by experiment. We plan to do this over the next year or so as part of S. Casey's PhD project, using developments of the new software above.

The series of eVLBI tests carried out over the last few years have clearly shown that use of the internet for VLBI is feasible, and that useful scientific results can be obtained. Our aim to to increase the bandwidth available for VLBI astronomers nearer to the limits set by current network and computer technology Continuing experiments will show where bottlenecks occur. It is already clear that the UKLight switched light path will give improved performance over the production network, even though a full eVLBI test has yet to be performed.

Ultimately, the European eVLBI community would like to transfer data in production mode, operating 24/7 for several weeks. Extended experiments using the UKLight infrastructure will be an important staging point in validating the use of switched-lightpath circuits for this purpose, as well as developing expertise in the eVLBI community.

Acknowledgements

The authors thank Steve Parsley, Arpad Szomoru and and Paul Burgess for their valuable contribution to running the experiments at JIVE and Jodrell Bank. Special thanks are due to Adam Mathews and Stephen O’Toole for making their project so successful.

References

[1] Thompson, A. R., Moran, J. P. and Swenson G. W., *Interferometry and Synthesis in Radio Astronomy* (2nd Ed) Wiley, 2001.
[2] Alef, W., in *Proceedings of the 7th European VLBI Network Symposium*, Toledo, p 237, October 2004, http://www.oan.es/evn2004/
[3] Born, M., and Wolf, E., *Principles of Optics*, 6th edition, p 508, Pergamon, 1980
[4] Ryle, M., and Hewish, A., Mem. R. Astro Soc. 67, 97, 1955
[5] Hughes-Jones, R., Parsley, S. and Spencer, R., Fut. Gen Comp. Sys. 19, 883, 2003
[6] http://www.evlbi.org/
[7] Hughes-Jones, R., Clarke, P., Dalison, S. and Fairey, G.. 2004, Fut. Gen. Comp. Sys., 21, 469, 2005
[8] http://www.hep.man.ac.uk/u/rich/
[9] Spencer, R., Hughes-Jones, Mathews, A., and O’Toole, S., in *Proceedings of the 7th European VLBI Network Symposium*, Toledo, p 261, October 2004, http://www.oan.es/evn2004/
[10] Picinbono, R., *Random Signals and Systems,* Prentice Hall, 1993
[11] Crovella, M. E., and Bestavros, A., , Proc. ACM Sigmetrics, May 1996
[12] Park, K., and Willinger, W., *Self-Similar Network Traffic and Performance Evaluation*, Wiley, 2000
[13] http://www.haystack.mit.edu/e-vlbi/
[14] http://www.eslea.uklight.ac.uk/
[15] http://www.jive.nl/