EXPReS/FABRIC Strategic Document:

Protocol Investigation for e-VLBI Data Transfer

Document FABRIC-1.2.1.001v1 for EU Project Number 026642*

Matthew Strong[1], Ralph Spencer[1], Richard Hughes-Jones[2], Simon Casey[1]

1 **Jodrell Bank Observatory, The University of Manchester, UK**
2 **Schuster Building, The University of Manchester, UK**

3 May 2005

* This work has received financial support under the EU FP6 Integrated Infrastructure Initiative contract number 026642, EXPReS.

CONTENTS

Abstract

This paper is a strategic document for the EXPReS/FABRIC project. The report outlines the developments and decisions required for choosing the appropriate IP protocol for use in e-VLBI. After introducing e-VLBI and its requirements, the various protocols in use on the Internet are discussed. VLBI has continuously streamed data, where individual packets are not particularly valuable. Maintenance of the data rate is important and so the requirements are quite different to those of e.g. file transfer where bit-wise correct transmission is required. The report discussed the actions required in order to make an informed decision and to implement a suitable protocol or protocols in the European VLBI Network.

1.0 Introduction

The e-VLBI project[1] is an upgrade to the current VLBI system. VLBI (Very Long Baseline Interferometry) is an aperture synthesis technique that utilizes radio telescopes from around Europe (and the world for global VLBI) to combine astronomical data in order to achieve high angular resolution observations. EVN (European VLBI Network) is the organization that administers VLBI operations primarily within Europe, but also with telescopes in Asia and South Africa. The EVN members operate 18 individual antennae; including some of the largest and most sensitive radio telescopes is the world. The telescopes observe the same cosmic radio source simultaneously, and currently the data are recorded on magnetic tapes or disk packs. These magnetic tapes or disk packs are then replayed and combined at the correlator, and data processed.

The e-EVN project is an upgrade to current EVN system that will give a real-time radio telescope as large as Europe with sub-microJansky sensitivity. Ultimately the aim of the project is to connect radio telescopes around Europe (and the world) with broad-band optical fibre networks to achieve real-time data transmission at rates of between 1 and 10 GBit/s. Data transmission rates of better than 1 GBit/s will permit noise levels within the observations that are better than 1 microJansky. Initial, proof of concept work, performed as a collaboration with the National Research Networks and Dante, and demonstrated at IGRID2002 [2] led to a collaboration of the EVN with that of the ESLEA[3] (Exploitation of Switched-Lightpaths for E-science Applications) to develop and test the concepts. Now further development is being made with the European Union funded EXPReS[4] project, starting in March 2006. This document is a deliverable for the FABRIC JRA project.

In order to achieve a real-time e-EVN interferometer the data taken at the radio telescopes must be transferred to the correlator in real-time, and not recorded and played back at a later date. This is achieved by the use of optical fibre networks. Optical fibres have a high bandwidth, and as such can transport large amounts of data between sites all across the world. There are two types of optical fibre networks being researched for use with e-VLBI data; these are the standard academic production internet and dedicated optical lightpath networks, often called OPN.

In order for these optical fibre networks to perform at their optimum, the right choice of transmission protocol must be made. This choice of protocol can have dramatic effects on the data transmission over the network, and so it is essential that the right choice is made. There are a number of well known transmission protocols that have been used much before, that can be used for e-VLBI, but there are other less well known protocols that could provide better data transmission for the specific goal of e-VLBI data transfer. The purpose of this report is to discuss the viable transmission protocols for e-VLBI, including their relative advantages and disadvantages and suggest which protocols should be investigated further, for future development of the e-VLBI data transfer system.

2.0 Basic Needs of VLBI systems

Radio astronomy data are inherently streamed. Signals from celestial sources are amplified, filtered and digitally sampled continuously. Full spectral information is still available even with single-bit sampling as was shown by van Fleck in the 1940's. Most systems use 2 or more bits giving a slightly higher signal to noise, and giving a higher dynamic range to cope with interfering signals from man made sources. A crucial factor in radio astronomy is that signals are averaged, in VLBI this can be effectively for 12 hours or more. The signal to noise ratio is proportional to $\sqrt{B\tau}$ where B is the bandwidth and τ the integration time, and all state-of-the-art observations are noise limited. Bandwidths of several hundred MHz are now in common use in modern telescopes. Multi-bit sampling at the Nyquist rate can therefore give data rates of 1 Gbps or higher, for example eMERLIN will use 2 GHz bands in 2 polarisations with 3 bits giving a data rate of 24 Gbps, after formatting and encoding this becomes 30 Gbps. Rates in VLBI are more restricted due to the need for data recording of the raw data, a problem not encountered by connected element arrays like eMERLIN. However it is the long term aim to achieve similarly high data rates by using the Internet. In Interferometry the signals needs to be brought together at a central site to be cross-correlated (effectively multiplied together in pairs and averaged). In EVN this correlator is situated at JIVE, Dwingeloo.

Current VLBI systems were designed around legacy tape recorder techniques. Multiple sampled digital data streams have headers containing timing and ancillary data added in a formatter before being sent to individual magnetic heads in the tape recorder. In the new disk based Mk5 systems these 'track' data are sent via a StreamStore card on to a disk. The PCEVN system being developed further elsewhere in FABRIC can also write to disk. Alternatively the data can be sent as packets to the internet in either system. Signal conditioning (amplification and filtering) prior to digitisation is performed by analogue electronics, though it is planned to replace these by digital systems in future. The current Mk5A system can record at up to 1024 Mbps, i.e. higher than available in 1 GE. The samplers and formatters are configured to work at data rates which change by factors of 2, so the next lowest rate is 512 Mbps.

The Mk5 and PCEVN allow continuous data streams to be sent over the internet, making use of facilities provided in the Linux kernel to packetise and format the data, including the necessary IP. TCP/IP is currently used, however this has inherent limitations as discussed later in this paper. Gigabit Ethernet network interface cards are readily available and these are use to put data on to LANs. Tests are underway to develop real-time capability at 512 Mbps data rate in

European e-VLBI, though to date reliable operation with the 6 connected telescopes is at 128 Mpbs. It is the aim of FABRIC to make the best use of the Internet bandwidth available, and so 512 Mbps is a reasonable aim given single 1 GE links to each telescope. However 10 GE is becoming more common, with the possibility of switched lightpaths being provide by GEANT2 and the NRENs, and so tests at 4 Gbps are envisaged as part of the FABRIC programme.

2.1 Value of e-VLBI data packets

The value of individual data bits as they traverse the internet highway is not particularly high. The signals are essentially Gaussian distributed random noise; one bit is much the same as any other. Loss of packets has the same effect as loss of observing time due to necessary telescope slewing movements on to source, time outs for calibration, source rising and setting etc. as long as lost packets are logged. The net effect is a loss of signal to noise, and a decrease of up to 5% or perhaps even more would not be noticed, though factors of 2 (resulting in a root 2 decrease in signal to noise from the equation above) would be. The maintenance of the data rate is however more crucial as the correlator needs to remain in synchronisation in order to line all the data from the various telescopes together.

These properties reflect in the requirements for any protocol which is used to transmit data. For example, the factor of 2 back-off in rate for TCP when a single packet is lost has a catastrophic effect, resulting in a need for re-synchronisation, and a loss of 41 % in signal to noise. In fact ~5 % of the packets could be lost (provided synchronisation of the correlator is maintained) before having a significant effect on the quality of the data, provided the protocol does not back off. This situation is quite different for most applications using IP where such error rates would be totally intolerable, and therefore e-VLBI has unique requirements.

3.0 Requirements for e-VLBI data transmission

It is important to understand what requirements e-VLBI data transmission puts on any protocol. e-VLBI requirements are very specific, and they ultimately determine whether a transmission protocol is suitable or not. In addition to this, it is desirable that the protocol should leave enough flexibility for further extensions to the system in the future, such that this exercise would not have to be repeated as the system develops.

So what are the requirements of a transmission systems and protocol for use with e-VLBI?

1. *Fast transport (>0.5 Gbps)* through long, large capacity optical networks.

2. *Fairly reliable* transmission. Although e-VLBI can operate with some packet loss, this is thought to be at the 2% level from theoretical calculations for the current systems where headers could be lost[5]. Thus, any transmission protocol would have to ensure reliable data transport losing less than 2% of packets in order for the correlator to remain in synchronisation. If the correlator can stay in synchronisation then higher loss rates can be tolerated (see section 2.1)
3. *Controlled data rates* – network transfer rates controlled by the application
4. *Bounded latency* – one way delay for link should remain constant ensuring timely delivery of data (jitter), inter-gap variation
5. *Bit wise correct data* – data which arrive should be bit wise correct
6. *Lost packets* – must be detected by the receiver and reported to the application
7. *Low re-ordering* – such that data can be placed in the correct order in a simple manner in the receiver
8. *Packet duplication* – must be minimal
9. *Distributed correlation* – where short sections of narrow band data (low data rate) are sent to many processors

The use of distributed correlation in a GRID type system is being investigated in FABRIC. The advantages and practicalities of such a system need to be determined, and a full specification of data transmission requirements obtained. For such a system, interfacing real time flows into grid environments may need the ability to send copies of data to multiple processor centres.

We note that the networks may consist of a mixture of switched light paths and production academic packet switched links. Optimum solutions for the two will have to be examined.

We also note that the information security policies of some networks may place restrictions on what kind of protocols can be used and on connectivity, and so discussions with network providers may be needed. So far we have had excellent cooperation with local and national research networks and also including DANTE and we expect to build upon these good relations.

3.1 Future developments

There are many new technological developments that are being investigated and implemented within the next few years. Systems such as MkVB, PCEVN2, eMERLIN-in where data flows into the eMERLIN correlator (in FABRIC at 4 Gbps), and eMERLIN-out where data flows out to JIVE (in EXPReS) are being designed now and have different data formats to those of current systems. The designs may result in special requirements for network data transmission which will need to be taken into account for the e-VLBI protocol decisions.

4.0 Transmission Protocols for e-VLBI

Data is transported around the world on networks using a series of protocols. These protocols allow the transmission of data from one point on the network to another. There are many protocols involved in many different layers of transmission of data on networks and as such we refer to a transmission stack. A typical transmission stack for a normal packet switched network is shown in figure 1 and broadly consists of application layer, a transport layer, a network layer and a data layer.

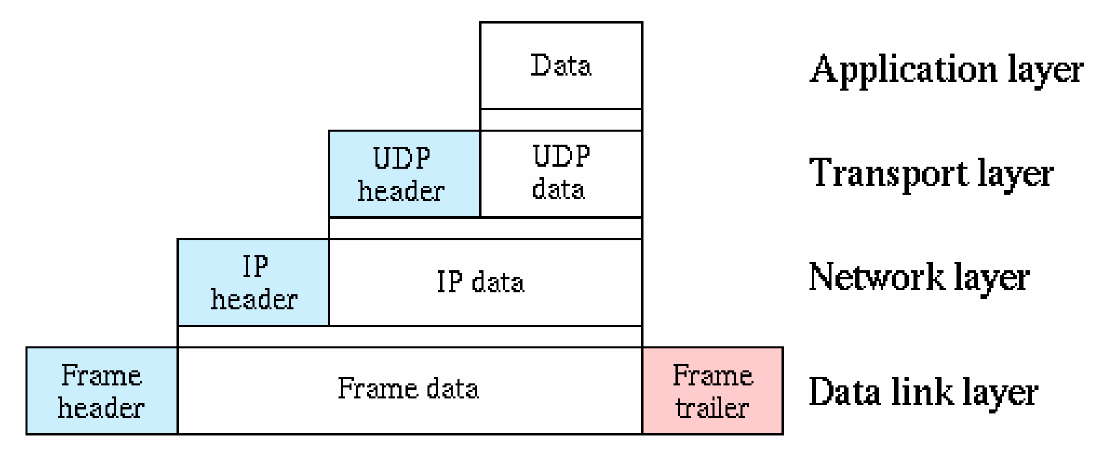

Figure 1 - Schematic representation of the protocol stack. Here UDP represents the transmission protocol.

There are a number of transport and application protocols that can be used with the e-VLBI system, all giving different transmission properties on the network. These protocols range from established protocols such as TCP and UDP, to less established ones such as SCTP and application protocols such as Tsunami. A list of the transport and application protocols examined in this report is given below.

Transport Protocols

- TCP
- UDP
- DCCP
- SCTP

Application Protocols

- Tsunami
- UDT
- RDMA
- Protocol off-load engines
- VSI-E (uses the RTP transmission protocol)

4.1 Transmission Control Protocol (TCP)

TCP guarantees reliable, in-order, and non-duplicated delivery data from sender to receiver. It is a byte-stream transport layer protocol that sits above the unreliable Internet Protocol (IP) packet based service. To use TCP, applications first create connections between themselves and then exchange data. On receipt of data from the application, TCP first divides this byte-stream into approximately evenly sized segments of information, assigns each segment a sequence number which is the number of the first byte of the application byte-stream in this segment, and then passes these segments to the IP layer. The size of a segment is determined by the Maximum Transmission Unit (MTU). The TCP module at the destination sends an acknowledgment (ACK) indicating the next byte expected for each received TCP packet. If an ACK is not received by the sender in a reasonable round trip time (RTT), indicating a timeout, or more than three ACKs are received with the same expected next byte number, indicating packet loss, then the corresponding segment is resent, hence allowing reliable data transmission. To achieve efficient reliable data exchange of the byte-stream, TCP uses a sliding window over the data with the reception of ACKs moving the window along, allowing more data to be sent.

However, this reliable data transmission comes at a price. If a packet has been lost, TCP interprets this as congestion on the network, and reduces the size of the sliding window, correspondingly dropping the rate at which packets are sent. TCP then increases the transmission rate slowly, and if there are no more lost packets, the transmission rate can reach a maximum value again after a period of time [6]. In fact, the recovery time can be long for some TCP transmission: about a minute for European transmissions and up to an hour for transatlantic transmission [5].

The long recovery time due to the TCP congestion control algorithm is an undesirable effect regarding the transmission of radio astronomical data, as VLBI data is generated at a constant rate and needs to be sent over the network at this rate if real-time VLBI is to be accomplished. In addition to this, VLBI is not greatly affected by the loss of a little information as the raw data is Gaussian noise. However, the use of TCP is currently common in network transmission with much software designed to use it. Therefore, TCP continues as a viable transmission protocol in the e-VLBI community even though it does not have optimum performance.

We note that recent advanced TCP stacks (such as FastTCP, HSTCP, Scaleable TCP, Hamilton etc.) use different congestion avoidance algorithms resulting in faster recovery and may be useful for e-VLBI.

4.2 User Datagram Protocol

UDP is another transport layer protocol originally designed by the US department of defence for use with the IP network layer protocol[7]. It provides a best effort datagram service between end systems. It is an unreliable service that provides no guarantees that delivery has occurred or no protection from duplication. A computer may send a UDP packet without first establishing a connection to the recipient. The UDP packets are sent over the network to a recipient host, listening for UDP packets with a particular port number. UDP does not know if packets are lost and thus cannot reorder packets. There is no congestion control in the UDP transport protocol.

For VLBI, a low level of missing packets is not necessarily a problem as the loss of such information only leads to a proportionate deterioration in signal-to-noise ratio. Also, because the protocol does not alter the transmission rate of the data, the rate is only determined by the available bandwidth of the network and hence this reduces problems with the transmission rate dropping below necessary e-VLBI rates.

4.3 The DCCP protocol

DCCP (Datagram Congestion Control Protocol) is a message-orientated transport layer protocol and is generally used in applications with timing constraints on the delivery of data[8]. Such applications include video streaming and internet applications such as internet telephony for example. The primary motivation for the development of DCCP is to provide a way for such applications to be able to use with variation congestion control mechanisms without having to implement them at the application layer.

DCCP is intended for applications that wish to be responsive to changes in network conditions and this is achieved by the use of flow-based congestion control algorithms. Several different congestion control algorithms are being developed including the TCP flow control. Note that DCCP does NOT guarantee reliable transmission and in order delivery.

At the current time, most applications that could benefit from the use of DCCP either use the TCP protocol, with its problems, or use UDP and implement their own congestion control algorithms (or use no congestion control at all). DCCP allows the user access to the congestion control systems, such that the user can set up the system for optimum use with its needs.

4.4 The SCTP protocol

SCTP (Stream Control Transmission Protocol) is a transmission protocol offering acknowledge, error-free, non-duplicate transmission of datagrams[9][10]. The difference between SCTP and the archetypical transmission protocol, TCP, is that multi-homing and the concept of several streams within a connection (or association). Where, in TCP, a stream is referred to as a series of bytes, in SCTP, a stream is referred to as a series of messages.

SCTP is a unicast protocol supporting transmission between two endpoints, although these endpoints can be multiple IP addresses. The SCTP transmission rate is adaptive, similar to TCP, and will scale back transfer to the prevailing load conditions of the network, and is designed to behave cooperatively with TCP sessions attempting to use the same bandwidth.

The multi-streaming function of SCTP allows data to be partitioned into multiple streams of data, each initially independent of the others ensuring that message loss in any one of the streams will only affect that particular stream, and not the others. SCTP accomplishes its multi-streaming function by creating independence between data transmission and delivery. This independence allows the receiver to determine immediately when a gap in the transmission sequence occurs, and whether the messages received following the gap are within the affected stream. The receiver can then continue to deliver messages to the unaffected stream, while buffering messages to the affected stream until retransmission occurs.

Another core feature of the SCTP protocol is its multi-homing feature. This is the ability of each endpoint of the connection to support multiple IP addresses. The benefit of this is potentially survivability of the session in the presence of network failure. This feature can provide effective solutions for failures in local LANs, as well as in the core network. Using multi-homed SCTP, redundant LANs can be used to reinforce the local access, while various options are available in the core network to reduce dependency of failures for difference addresses.

4.5 The Tsunami protocol

Tsunami is an application level file transfer protocol that uses UDP as the transport mechanism to transfer files[11]. Its architecture is similar to FTP in that it uses a control session to authenticate and negotiate the connection, and a data session to transfer the target file. The Tsunami architecture follows a usual client-server model. The client asks the server for a file using the typical TCP control port. The server forks a thread to handle this request, and then goes back and waits for the next connection. The forked thread checks for the file and attempts to transfer it using a known UDP port. This transfer then commences in blocks, where the block size is variable.

UDP does not have guarantee reliable, in order delivery, and so Tsunami has its own retransmission and reordering mechanisms. If any block of data is lost, delayed or out of order, a request for retransmission is sent on the TCP control port.

4.6 UDP Based Data Transport Protocol (UDT)

UDP is an application layer data transport that uses UDP to transfer bulk data and has its own reliability control and congestion control algorithms. It is designed for use with emerging distributed data intensive applications for use over wide area high speed (>1Gbit/s) networks, both private and shared[12].

4.7 Remote Direct Memory Access (RDMA)

RDMA is a process where two or more computers communicate via direct memory access directly from the main memory of one system to the main memory of the others. For this process, there is no CPU, cache or context switching overhead needed to perform the transfer, and such transfers can be carried out in parallel with other system operations[13].

It can be seen that this could be useful for e-VLBI, as such a process is well suited to applications where high throughput, low latency networking is necessary.

4.8 Protocol off-load engines

Protocol offload engines are a technology designed to improve the performance[14] of a protocol, specifically by moving the processing from the main CPU to a separate dedicated subsystem.

With modern network interface cards (NIC) the TCP/IP checksum calculation and data segmentation into MTU-sized chunks, known as TCP Large Send Offload (LSO), or TCP Segmentation Offload (TSO), can be performed in the NIC, freeing the main CPU from these computations. A TCP Offload Engine (TOE) goes further by performing all the TCP protocol processing on the TOE Card, including ACK processing and re-transmissions. This can significantly reduce the number of interrupts that the main CPU has to process, allowing more CPU power to be given to the application.

Thus, it can be seen that protocol offload engines can create a performance improvement in data transmission, with current devices capable of line speed performance on 10 GE. However they tend only to offer standard TCP congestion algorithms and hence not give the improved recovery times of the advanced TCP stacks discussed above.

4.9 The VSI-E protocol

VSIE (VLBI Standard Interface – Electronic) is a specific method of e-VLBI data transmission over global networks[15][16]. It makes use of the Real-time Transport Protocol (RTP) suite and has four main specifications:

1. *Interoperability* – through a common data format.
2. *Internet Friendliness* – through the use of protocols that are we known and commonly used throughout the internet community.
3. *Ease of implementation* – through the use of existing or new libraries.
4. *Transport flexibility* – through the use of a framework that will allow users to choose their transport mechanism/protocol to suit their network and/or throughput requirements.

These specifications are met through the use of the RTP protocol suite, which specifically includes:

1. *RTP* – provides a means for encapsulating real-time data streams and transporting them across Wide Area Networks.
2. *The Real-time Transport Control Protocol (RCTP)* – which provides a control channel for RTP streams that is used to exchange management information as well as sender/receiver-side statistics and timing synchronisation information.

In the design and development of VSI-E, the developers have a wealth of information to call on, due to the fact that RTP is an established system, used throughout the internet community for many years. RTP is considered ‘internet friendly’ within the internet community; a property which can easily be extended to accommodate e-VLBI requirements.

5.0 Discussion: The Choice of Transmission Protocol?

The choice of a transmission protocol for use with e-VLBI is a complex issue. e-VLBI requires fast, fairly reliable transmission over long network links with large round trip times. It would seem logical to investigate the use of established protocols such as TCP and UDP initially, as there is wealth of experience in these, and such protocols might easily be tailored to e-VLBI’s needs. However, it also seems logical to explore other, newer transmission systems, such as Tsunami, as such a system could fit e-VLBI specification without much modification.

Currently, EVN tests use TCP in the transmission of e-VLBI data, with some success. However, it is well known that the congestion control algorithm within

TCP is not suited to e-VLBI’s needs. Indeed, packet loss within the network is not usually at a high level, but packet loss in the end hosts can be. The TCP congestion control algorithm tries to determine if congestion is present by observing packet loss, and then reduces the rate of data transmission to allow “fair sharing” of the network between users. Thus, without changes to the flow control within TCP, it will not be able to meet the e-VLBI specification fully.

Using the pure UDP transmission protocol is another viable option. UDP does not possess any congestion controls, or measures for reliable transfer. Hence, in terms of e-VLBI, it will just transmit the telescope data continuously at the maximum bandwidth the link will allow. However, any packet loss in the end hosts or network will result in loss in the pre-correlated data streams. How much loss can be tolerated before correlation is no longer possible is a source of interest at the moment, with tests planned for July 2006 to quantify this. Currently, estimates of approximately 5 % loss are thought to be tolerable, although the tests will show whether this is a good approximation or not. e-VLBI UDP transmission software is currently being developed and tested by Richard Hughes-Jones and Simon Casey at the University of Manchester with encouraging results. In initial memory to memory transmission tests, transmission rates of in excess of 800 MBit/s have been obtained, with zero packet loss. It seems that such a system is worth pursuing in the development of e-VLBI data transmission.

Whilst the continuing development of established protocols such as UDP is necessary, it is also evident that some of the newer, more specific transmission programs require further investigation to establish if they are useful in the development of e-VLBI. Certainly, transmission systems such as Tsunami appear to be very relevant in the transmission of e-VLBI data. Tsunami itself is one of a number of new data moving application protocols, developed in 2002 – 2003, and is capable of tolerating a reasonable amount of packet loss, somewhere between 1 and 7 %. Even the first e-VLBI experiments with Tsunami have been very successful, with file-to-file transfer speeds of 640 Mbit/s achieved between Espoo, Finland and Dwingeloo, the Netherlands, and also 400 Mbit/s between Kashima, Japan and Dwingeloo. In comparison, the TCP transmission speeds were 10 and 2 MBit/s respectively over the same two links, under the exact same conditions. Thus it can be seen that the potential of such transmission systems as Tsunami is vast, and could be a great advantage in the development of e-VLBI.

One of the current issues being investigated for the use with e-VLBI, is the development of a distributed software correlator. Part of this investigation is the possibility of splitting the observing band into smaller frequency-time chunks, sending these chunks to different stations to be correlated, and then adding the correlation results back together later in the process. Thus, for this to be successful, the e-VLBI transmission protocol would have to have the ability to send specific data to more than one destination. This, referred to here as multi-homing, is a specific task not common in all transmission systems, with protocols such as SCTP having this capability. Indeed, further investigation into protocols such as SCTP is necessary to ascertain whether such protocols could be of use with e-VLBI.

5.1 Further Investigations Needed

It is clear that it is not known at the present time what is the optimum transport or application protocol for e-VLBI. Much more investigation is needed in order to make a more consider choice. In particular we note the following areas of possible investigation:

1) TCP - we already aware of problems, further tests are needed
2) UDP - tests just started using VLBI-UDP, more data to come
3) Tsunami - members of the group have made initial tests
4) DCCP - being developed in the ESLEA projects (UCL and Manchester), plans are to participate in porting to vlbi and further development
5) TOE - offload engines may reduce CPU load and need to be investigated further
6) VSI-E - being developed by Haystack, needs looking at in detail
7) Other protocols and TCP variants need examination
8) Distributed correlation – further discussions within FABRIC needed
9) FPGA implementation (eMERLIN) – transmitting VLBI data directly to the network from a hardware card, the effectiveness and flexibility needs investigation

5.2 Actions

The following outline plan of action is needed for us to be able to come to an optimum solution, ready in time for full implementation in EVN for the full data rate tests envisaged near the end of the project.

- Take part in planned tests on currently implemented protocols
- Investigate properties of other protocols in detail including GRID suitable protocols (paper exercise)
- Chose favourable candidates
- Implement candidates where possible given resources available (include discussion with other groups engaged in network research)
- Tests on favourable candidates using dummy and VLBI data
- Investigate GRID suitable protocols in collaborations with WP 2
- Chose optimum protocol(s) and suitable implementation

- Implement in EVN
- Evaluation of FPGA generated transmissions on both Private Optical Networks and the standard Academic internet.
- Write up reports

6.0 Conclusions

It is evident from the discussion previous that much work is necessary on determining the best transmission protocol for use with e-VLBI. The final choice can only be made when the choice of correlation system is decided as multi-homing may have to be considered. However, from the discussion above it is evident that the current TCP system does not make optimal use of the system, mainly due to its congestion control algorithm. Thus, it seems logical to move on from a TCP based system, and use a more appropriate protocol.

There is much investigation needed, both with the UDP transport protocol and UDP systems (such as Tsunami) to provide a valid alternative to TCP. Indeed, tests with the VLBI_udp software already made show encouraging results – with transfer speeds of in excess of 800 MBit/s. Also, initial tests with Tsunami are also encouraging, showing in excess of 50 and 100 times better transmission speeds than TCP under the same conditions. Protocols such as *VSI-E* could provide a flexible alternative to TCP, with the ability to choose the transmission protocol within the transmission system.

From the discussion presented here, it is important that the choice of transmission protocol is correct, and the wrong choice could lead to poor performance. Much testing and investigation is needed in order to ensure this decision is the correct one.

References

[1]EVN e-VLBI homepage, *http://www.e-VLBI.org/e-VLBI/e-VLBI.html*
[2]Hughes-Jones, R., Parsley, S., Spencer, R., 2003
"High Data Rate Transmission in High Resolution Radio Astronomy - vlbiGRID",
FGCS Special Issue: IGRID2002 Vol 19 (2003) No 6
[3]Exploitation of Switched-Lightpaths for E-Science Applications, homepage
http://www.eslea.uklight.ac.uk/
[4]EXPReS project proposal document
[5] Spencer. R., Hughes-Jones. R., Matthews. A., O'Toole. S., 2004.,
Proceedings of the 7th European VLBI Network Symposium
[6]TCP RFC 793 webpage, *http://www.faqs.org/rfcs/rfc793.html*
[7] UDP RFC 768 webpage, *http://www.ietf.org/rfc/rfc768.tx*t

[8] DCCP wikipedia page, *http://en.wikipedia.org/wiki/Datagram_Congestion_Control_Protocol*
[9]SCTP for beginners, *http://tdrwww.exp-math.uni-essen.de/inhalt/forschung/sctp_fb/*
[10]SCTP RFC2960, *http://www.ietf.org/rfc/rfc2960.txt*
[11]Tsunami – A Study, *http://www-iepm.slac.stanford.edu/bw/Tsunami.htm*
[12]UDT source - *http://udt.sourceforge.net/*
[13] NetworkWorld web resource - *http://www.networkworld.com/details/5221.html*
[14] TCP offload engine Wikipedia page - *http://en.wikipedia.org/wiki/TCP_Offload_Engine*
[15]VSI-E Software Suite, Lapsley. D., Whitney. A., 2004, Proceedings of the 7th European VLBI Network Symposium
[16]VSI-E using the Real-time Transport Protocol, Draft Proposal for VSI-E – Rev 2.7, 2004