e-VLBI data transfer from Onsala and Metsähovi to the Bonn correlator

R. Haas
Chalmers University of Technology, Onsala Space Observatory, SE-439 92 Onsala, Sweden

J. Wagner, A. Mujunen, J. Ritakari
Helsinki University of Technology, Metsähovi Radio Observatory, FIN-02540 Kylmälä, Finland

A. Müskens, C. Dulfer, A. Bertarini
University Bonn, Institute for Geodesy and Geoinformation, Nußallee 17, DE-53115 Bonn, Germany
at Max Planck Institut for Radioastronomy, Auf dem Hügel 69, DE-53121 Bonn, Germany

Abstract. During autumn 2006 and spring 2007 we performed a series of e-VLBI data transfer experiments from the Onsala Space Observatory (Sweden) and the Metsähovi Radio Observatory (Finland) to the VLBI correlator at the Max Planck Institute for Radioastronomy in Bonn (Germany). We used the Tsunami protocol for data transfer both in off-line and real-time mode and successfully transferred VLBI data with data-rates of up to 800 Mbit/s (off-line) and 256 Mbit/s (real-time). Detailed comparisons of the e-VLBI transferred data with traditionally shipped data show that the e-VLBI data transfer is reliable. We describe a new strategy for operational e-VLBI data transfer.

Keywords. e-VLBI, optical fibre networks, Tsunami protocol, correlation

1 Introduction

The turn-around time for VLBI experiments in the IVS rapid series R1 and R4 today is still on the order of several days. One of the main reasons is that the VLBI observational data are still recorded on Mark5 modules that are later shipped via mail services to the correlators. Even inside Europe the shipment of Mark5 modules takes several days, e.g. about 3-4 days from the Onsala Space Observatory, Sweden, to the correlator at the Max Planck Institut for Radioastronomy in Bonn, Germany. To improve the timeliness of geodetic VLBI results, the time needed for the data transfer from the observing stations to the correlators needs to be reduced.

One attractive option is to use electronic data transfer via optical fibre networks, so-called e-VLBI data transfer. An operational use of e-VLBI data transfer would speed up the turn-around time for VLBI experiments and additionally avoid expenses for the shipment of Mark5 modules, both from the observing sites to the correlators, and back.

During the last couple of years several VLBI stations have been connected to high-speed optical fibre networks. Already in 2003 Onsala and Metsähovi were connected to high-speed optical fibre backbones, and recently also the Bonn correlator was connected. To exploit these existing high-speed connections we started in the fall of 2006 a series of test-experiment for e-VLBI data transfer from the VLBI stations Onsala and Metsähovi to the Bonn correlator. Different data transfer options were tested and we developed a strategy that can be used for routine application of e-VLBI data transfer. This work also prepares for near real-time VLBI observations, e.g. for Earth rotation observations.

2 Equipment and network for the e-VLBI data transfer tests

The two VLBI stations Onsala and Metsähovi are connected by 1 Gbit/s and 10 Gbit/s, respectively, to their national university networks SUNET (SUNET, 2007) and Funet (Funet, 2007). These are connected via the NORDUnet IP backbone network (NORDUnet, 2007) to the pan-European multi-gigabit data communications network GÉANT (GÉANT, 2007). The optical fibre from Onsala to Chalmers University of Technology is equipped with a coarse wavelength division multiplexer (CWDM) and has been described in Haas & Elgered (2006). The CWDM allows a shared use of the optical fibre with different wavelength, different MTU buffer sizes and different data rates, and efficiently separates the VLBI data transfer from other observatory traffic. Figure 1 shows schematically the connection from the VLBI computers at Onsala to the computers at the correlator at the Max Planck Institute for Radioastronomy in Bonn. A similar graph could of course also be shown for the connection between Mestähovi and Bonn but is omitted here in order to save space.

Onsala and Metsähovi are both equipped with

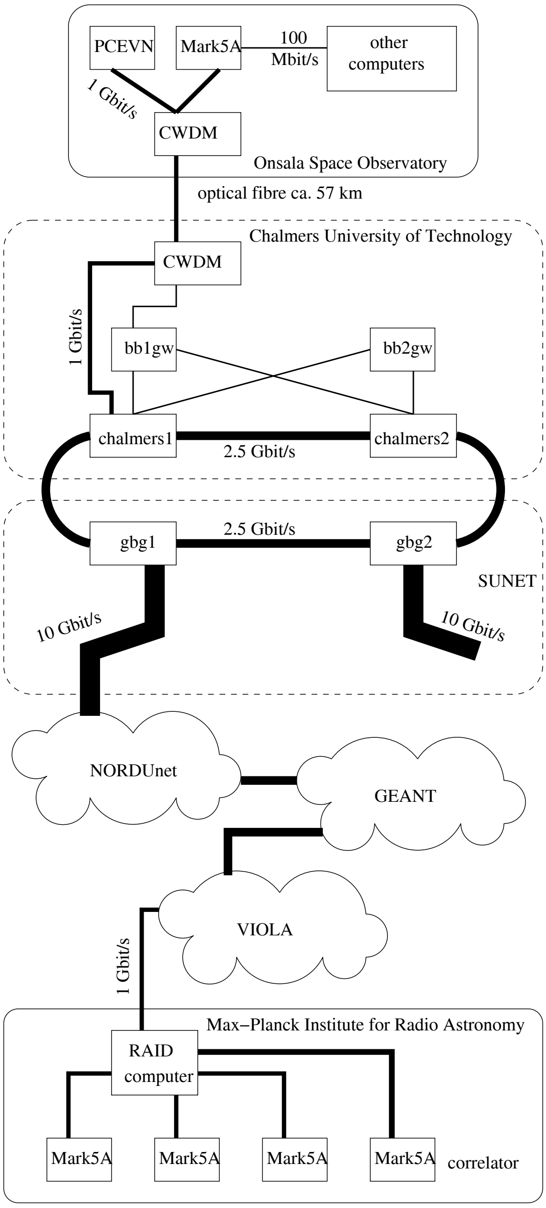

Figure 1. Schematic visualization of the connections from the VLBI computers at the Onsala Space Observatory to the computers of the correlator at the Max Planck Institute for Radioastronomy in Bonn.

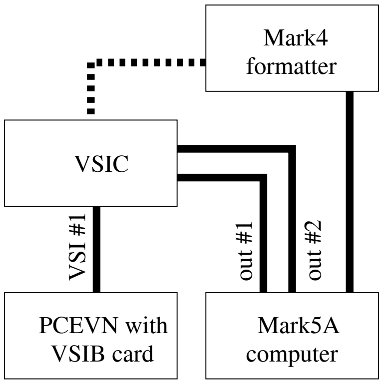

Figure 2. Connection of the PCEVN computer to the Mark4 formatter using the VSI converter.

Mark5 data acquisition systems and one (Onsala) or several (Metsähovi) PCEVN computers (Ritakari & Mujunen, 2002). The PCEVN at Onsala has a RAID system with a capacity of 1 TB, while the PCEVN computers at Metsähovi have RAID systems with capacities of up to 1.5 TB. For our e-VLBI data transfer tests the PCEVN computers were daisy-chained via a VSI-converter (Ritakari & Mujunen, 2002) to the Mark5A computers (solid lines in Fig. 2). This allows simultaneous parallel recording of VLBI data on both the Mark5 modules and the PCEVN RAID system, and real-time data-streaming with the PCEVN. The VSIC may be connected directly to the Mark4 formatter, too (dotted line in Fig. 2), thus bypassing the need for a Mark5 computer entirely.

Since mid 2006 the Bonn correlator has a 1 Gbit/s connection to the German research network VIOLA (VIOLA, 2007) that is connected to the European backbone network GÉANT. At Bonn there is a RAID-PC with a capacity of several TB, and several Mark5A computers for the actual correlator.

3 The Tsunami protocol for high-speed transfer of VLBI data

The Tsunami protocol (Tsunami UDP protocol, 2007) is an open-source protocol for data transfer that was initially developed by Indiana University and then successfully adapted by the Metsähovi group to support realtime and non-realtime VLBI applications. It combines Transmission Control Protocol (TCP) and User Datagram Protocol (UDP) data transfer to allow high-speed transfer of large amount of data.

Figure 3 schematically describes how the data transfer with the Tsunami protocol works. A low bandwidth TCP control connection establishes the connection between client and server, requests file transfer, controls the file transfer and feeds back the link quality. A high bandwidth UDP connection carries the actual user data, e.g. for our application the VLBI observational data. The client controls all aspects of the data transfer and determines whether data needs to be re-transmitted, or not. This makes the data transfer very fast and reliable. In off-line applications the client can control the protocol to be effectively loss-less. In real-time applications, there are timing and RAM buffer constraints that limit re-transmission effectiveness. Depending on the severity and duration of network congestion some data loss may, though need not, occur. Small amounts of data loss are often considered common and acceptable in real-time e-VLBI (Spencer et al., 2004).

Pre-requisites for an effective use of the Tsunami protocol for high-speed transfer of VLBI data are that the observatories are equipped with commodity PCEVN hardware with integrated 1 Gbit/s Ethernet and a VSIB PCI card. The correlator needs a fast receiving PC with a RAID system. The observatories and the correlator need of course to be connected to high-speed optical fiber networks. The computers on both sides (observatory and correlator) have to be time-synchronized by the Network Time Protocol (NTP, 2007) for real-time Tsunami applications.

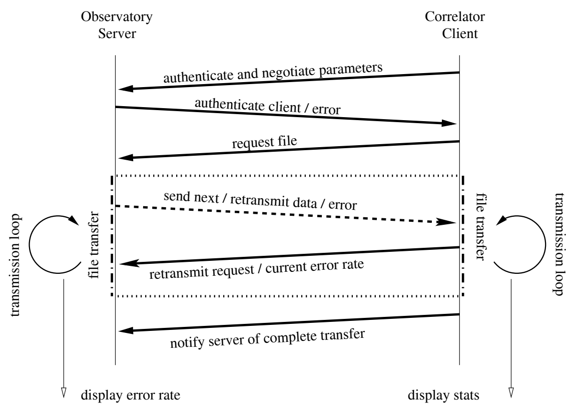

Figure 3. Schematic description of data transfer with the Tsunami protocol. See text for further explanations. Solid and dashed arrows depict TCP and UDP data transfer, respectively.

In off-line mode, a Tsunami server is started at the observatory, and a Tsunami client at the correlator. The Tsunami client connects to the server, starts the data transfer and receives the data that are stored on the RAID system of the sending PCEVN. The client can be called from scripts.

In real-time mode, a Tsunami client is started at the observatory. The server reads VSI bus data directly from the VSIB board of the PCEVN. The Tsunami client at the correlator requests experiment scans from the observatoriy using EVN Filenaming Convention compliant file names that contain auxiliary information about the desired scan. At the requested start time the server begins to stream the required amount of data to the client. The necessary Tsunami client scripts can be generated with a Metsähovi-provided snap-file transform script that reads the usual snap-files created when drudging the VLBI schedule files.

Data loss in real-time Tsunami applications can occur when the network link cannot reliably support the necessary data rates, or if the client PC is not capable of writing to the RAID system at the requested data rate. At high transfer speeds these effects can be more pronounced, but at slow speeds such as 128 and 256 Mbit/s they are usually not a problem. Short duration network congestion can be overcome by buffering of data in the RAM on the server side. The server is re-using the RAM ring buffer that is transparently provided by the VSIB card driver and has a typical size of 144 Mbytes.

4 Experience with e-VLBI data transfer to the Bonn correlator

During autumn 2006 and spring 2007 we performed a series of e-VLBI data transfer tests, see Table 1.

For the very first test we used the EGAE-software (EGAE, 2007) together with the Big Block File Tranfer Protocol (bbFTP, 2007), a file transfer software that is optimized for large files (> 2 GB), and transferred the Euro.84 data from Onsala to Bonn. The EGAE software first read out the data from the Mark5 module and saved them on the internal system hard-disk of the Mark5 computer. Then the data were transferred with bbFTP to the Bonn correlator. Once

Table 1. Experiments in fall 2006 and spring 2007 with e-VLBI data transfer to the Bonn correlator.

Session	data-transfer	station	
Euro.84	off-line EGAE-BBFTP	On	
Euro.85	off-line Tsunami	On	Mh*
T2.047	off-line Tsunami		Mh
T2.048	off-line Tsunami	On	
T2.049	off-line Tsunami	On	
Euro.87	off-line Tsunami		Mh
R1.258	real-time Tsunami	On	
R1.262	real-time Tsunami	On	
R1.263	real-time Tsunami	On	
R1.265	real-time Tsunami	On	

(*) partly real-time

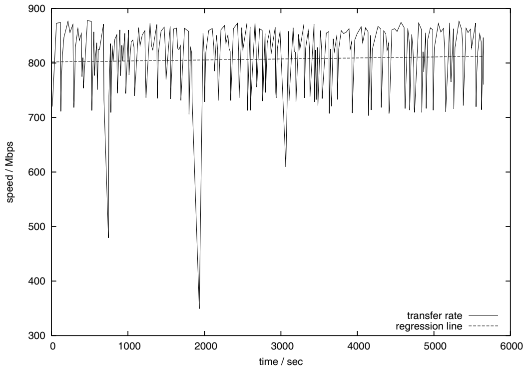

Figure 4. Euro.85 data transfer from Metsähovi to the Bonn correlator using the Tsunami protocol in off-line mode. The total amount of 572 GB were transferred in less than 1 hour and 40 minutes, with an average throughput of 806 Mbit/s.

a scan was successfully transferred, it was deleted from the system hard-disk of the Onsala Mark5 computer, and the next scan was read out from the Mark5 module. This procedure turned out to be restricted by the repeated read-write processes in connection with the data transfer from the Mark5 module to the system disk, and the bbFTP protocol. Thus, this test was quite slow and the effective throughput was only about 94.7 Mbit/s.

For the next tests we used the Tsunami protocol in off-line mode. Data were first recorded on the RAID system of the PCEVN computers at Onsala and/or Metsähovi, simultaneously to the data recordings on the Mark5 modules at the stations. For this purpose the PCEVN computers were daisy-chain connected to the Mark5A-computers, as described in Section 2. After the experiment the data were transferred with Tsunami protocol in off-line mode to the RAID-computer at the Bonn correlator. Data rates of up to 800 Mbit/s were achieved, see Fig. 4. Actually, some scans of Euro.85 were transferred with real-time Tsunami from Metsähovi, and the rest off-line.

After the off-line tests we tested also data transfer with the Tsunami protocol in real-time mode, from Onsala to Bonn. We tested this for a couple of R1-experiments that have a data rate of 256 Mbit/s. Additionally to the real-time streaming, we also recorded the data on the RAID system of the PCEVN computer at Onsala. However, due to capacity restrictions of currently 1 TB, we could only record every second scan on the PCEVN. With the real-time streaming we achieved average throughputs during the 24 hours of observations that were very close to the nominal data rates, see Fig. 5. Only a small amount of data lost due to congestion problems. The missing data were identified and then transferred after the experiment with off-line Tsunami from the PCEVN, or in case the missing data were not saved on the PCEVN, they were read out from the Mark5 module with the command 'disk2file' to the PCEVN and then transferred with off-line Tsunami.

The transferred data were first recorded on the RAID-computer at the Bonn correlator, and then transferred to Mark5 modules, using the

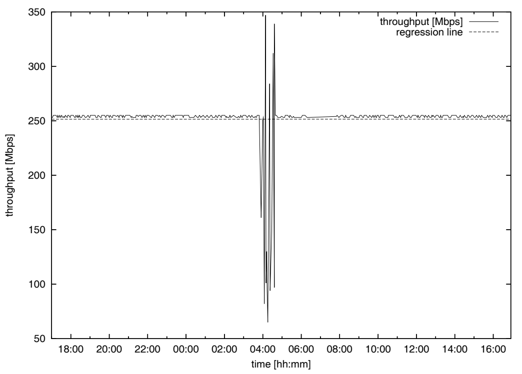

Figure 5. R1.265 real-time data transfer from Onsala to the Bonn correlator using the Tsunami protocol in real-time mode. The experiment had a data rate of 256 Mbit/s. Some data loss occured around 4:00 UT and an average throughput of 251 Mbit/s was achieved.

EGAE-software. For all experiments we also shipped the corresponding Mark5 modules to the Bonn correlator to allow a detailed investigation of the data transfer and comparison of the data. Figure 6 shows a comparison of scan-lengths from an autocorrelation of the two data sets. Positive values indicate that the Mark5-scans are longer, negative values indicate that the PCEVN-scans are longer. It appears that the PCEVN recorded, e-VLBI data transferred scans in general are slightly longer. The reason is that the recording on the PCEVN and the Tsunami transfer are independent from the Field System computer. PCEVN recording and transfer are triggered at the first 1 PPS and continue until the precise amount of requested scan data has been captured. The Mark5 recording in contrast lacks 1 PPS triggering and is dependend on the Field System computer communication and all station related procedures (e.g. preob) that are executed before the actual data recording on the Mark5 modules. These add a varying start-up delay to Mark5 recordings.

Figure 7 shows an example of a fringe plot obtained with e-VLBI transferred Onsala data for Euro.85.

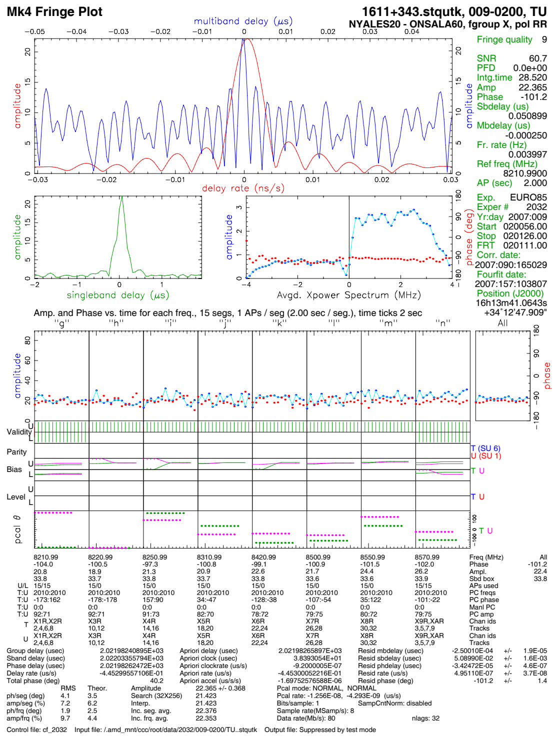

Figure 7. Fringe plot with e-VLBI transferred Onsala data for Euro.85.

5 New strategy for operational e-VLBI data transfer

Based on our experience with e-VLBI data transfer from Onsala and Metsähovi to the Bonn correlator using PCEVN and the Tsunami protocol, we propose a new strategy for operational e-VLBI data transfer. This strategy can and should be used for operational VLBI within the IVS and also prepares for possible near real-time correlation. The strategy involves:

1. real-time data transfer with the Tsunami protocol from the observing station to the correlator
2. simultaneous recording of the data on the PCEVN (backup-1)
3. recording of VLBI data on Mark5 modules (backup-2)

The correlator checks either at the end of the observing session or already during the session whether the real-time data transfer was successful and no data loss occurred. These checks of data completness can be done by scripts that e.g. compare the expected size of the files to be transferred and recorded at the correlator with the size of the actually recorded files at the correlator. Such scripts have been developed at the

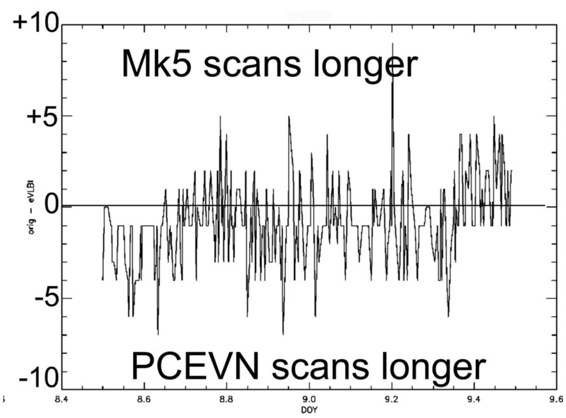

Figure 6. Comparison of scan-length from autocorrelation of data that were recorded on Mark5 modules and shipped to the Bonn correlator, and data that were transferred with the Tsunami protocol to the Bonn correlator.

Bonn correlator. In case that some data loss is detected, the incomplete or missing scans can be transferred off-line after the session with the Tsunami protocol directly from backup-1 on the PCEVN RAID system. In case there should be any problem with backup-1 on the PCEVN, there is backup-2 on the Mark5 module. In this case the data can be read out from the Mark5 module using the 'disk2file' command and then transferred off-line to the correlator. In the absolutely worst case that the network connection between the observing site and the correlator is completely broken and no data can be transferred electronically at all, the complete Mark5 module could be shipped traditionally via mail to the correlator. As soon as the correlation is completed, both backups can be removed, i.e. the files on the PCEVN RAID system can be deleted and the Mark5 module can be erased, re-conditioned and re-used for the next session.

6 Conclusions and outlook

The e-VLBI data transfer experiments performed in the fall of 2006 and the spring of 2007 with Onsala, Metsähovi and Bonn show that this approach to send geodetic VLBI data to a correlator works well. The data transfer, in particular with the Tsunami protocol, performed very well. Our tests show that it is possible to transfer a 24 h geodetic VLBI experiment to the correlator either in real-time or within two hours after the end of the experiment.

With standard 1 Gbit/s connections to the optical fibre backbone and standard Maximum Transmission Unit (MTU) size of 1500 a throughput of up to 800 Mbit/s was reached with off-line Tsunami. Real-time Tsunami worked reasonably well for several hours with data rates of 256 Mbit/s. Some data loss due to network congestion was still observable. However, we anticipate that this data loss can be avoided when using 10 Gbit/sec connections or light-paths.

Onsala is currently in a process of upgrading its network connection from 1 Gbit/s to 10 Gbit/s. Furthermore, does Onsala consider a possible upgrade of the capacity of the PCEVN RAID system from currently 1 TB to 2 TB. This will allow to record complete R1- and R4-experiments on the PCEVN. These R-experiments currently have typically 1.5 TB of observational data per station. Onsala aims at sending all VLBI data that is going to be correlated at the Bonn correlator via e-VLBI and not to ship any Mark5 modules to Bonn anymore.

Many VLBI stations in Europe and other parts of world have already today PCEVN equipment and are connected to high-speed optical fibre backbones. Thus, we propose to start the operational use of e-VLBI data transfer. A new strategy for an operational e-VLBI data transfer for the IVS sessions was described here.

Acknowledgment

We acknowledge that this work was partly funded by EXPReS. EXPReS is an Integrated Infrastructure Initiative (I3), funded under the European Commission's Sixth Framework Programme (FP6), contract number 026642 EXPReS.

References

bbFTP (Big Block File Tranfer Protocol), `http://doc.in2p3.fr/bbftp/`

EGAE (Experiment-Guided Adaptive Endpoint software), `http://evlbi.haystack.mit.edu/twiki/bin/view/EVLBI/EGAE`

Funet (Finnish University and Research Network), `http://www.funet.fi`

The GÉANT network, `http://www.geant.net`

Haas, R., and Elgered, G., 2006, The IVS network station at the Onsala Space Observatory, In: D. Behrend and K. D. Baver (eds.): *IVS 2005 Annual Report*, NASA/TP-2006-214136, 111–114.

NORDUnet IP Network `http://www.nordu.net`

NTP: The Network Time Protocol, `http://www.ntp.org/`

Ritakari, J., and Mujunen, A., 2002, A VSI-H Compatible Recording System for VLBI and e-VLBI, In: N.R. Vandenberg and K.D. Baver (eds.): *IVS 2002 General Meeting Proceedings*, NASA/CP-2002-210002, 128–131.

Spencer, R., Hughes-Jones, R., Mathews, A., and O'Toole, S., 2004, Packet Loss in High Data Rate Internet Data Transfer for eVLBI, In: R. Bachiller, F. Colomer, J.-F. Desmurs, P. de Vicente (eds.): *Proc. of the 7th European VLBI Network Symposium*, 261–264, `http://arxiv.org/abs/astro-ph/0501018`

SUNET (Swedish University Computer Network), `http://www.sunet.se`

Tsunami UDP protocol, `http://tsunami-udp.sourceforge.net/`

VIOLA (Vertically Integrated Optical Testbed for Large Applications in DFN), `http://www.viola-testbed.de/`