SA1: Production e-VLBI service

Arpad Szomoru
JIVE

Outline

- e-VLBI in practice
- Work packages
- Soft- and hardware developments

e-VLBI science/test runs

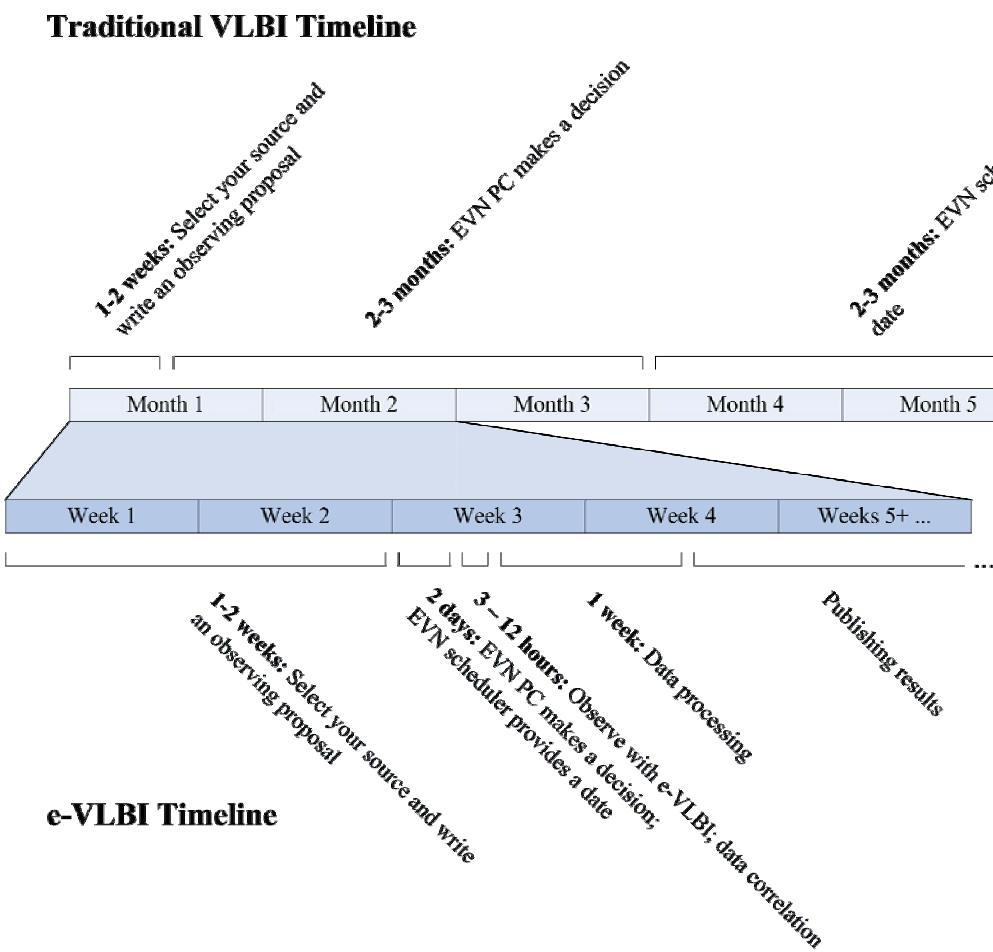
- 12 e-VLBI science projects accepted:
  - 2 failed
  - 3 Target of Opportunity (Cygnus X-3, GRS1915+105)
  - 3 determination of compactness of calibrator or target
  - 3 part of multi-wavelength campaign
  - 1 adaptive observation of 16 X-ray binaries (no detections..)

- Rapid access to EVN provides clear benefit to users (important for calibrator/multi-wavelength projects)
- Follow up of bursting transients only moderately successful
  - ❖two week delay between proposal deadline and actual observations is simply still too long

e-VLBI: operational improvements

- Overall improvement in first year:
  - Robustness
  - Reliability
  - Speed
  - Ease of operation
  - Station feedback

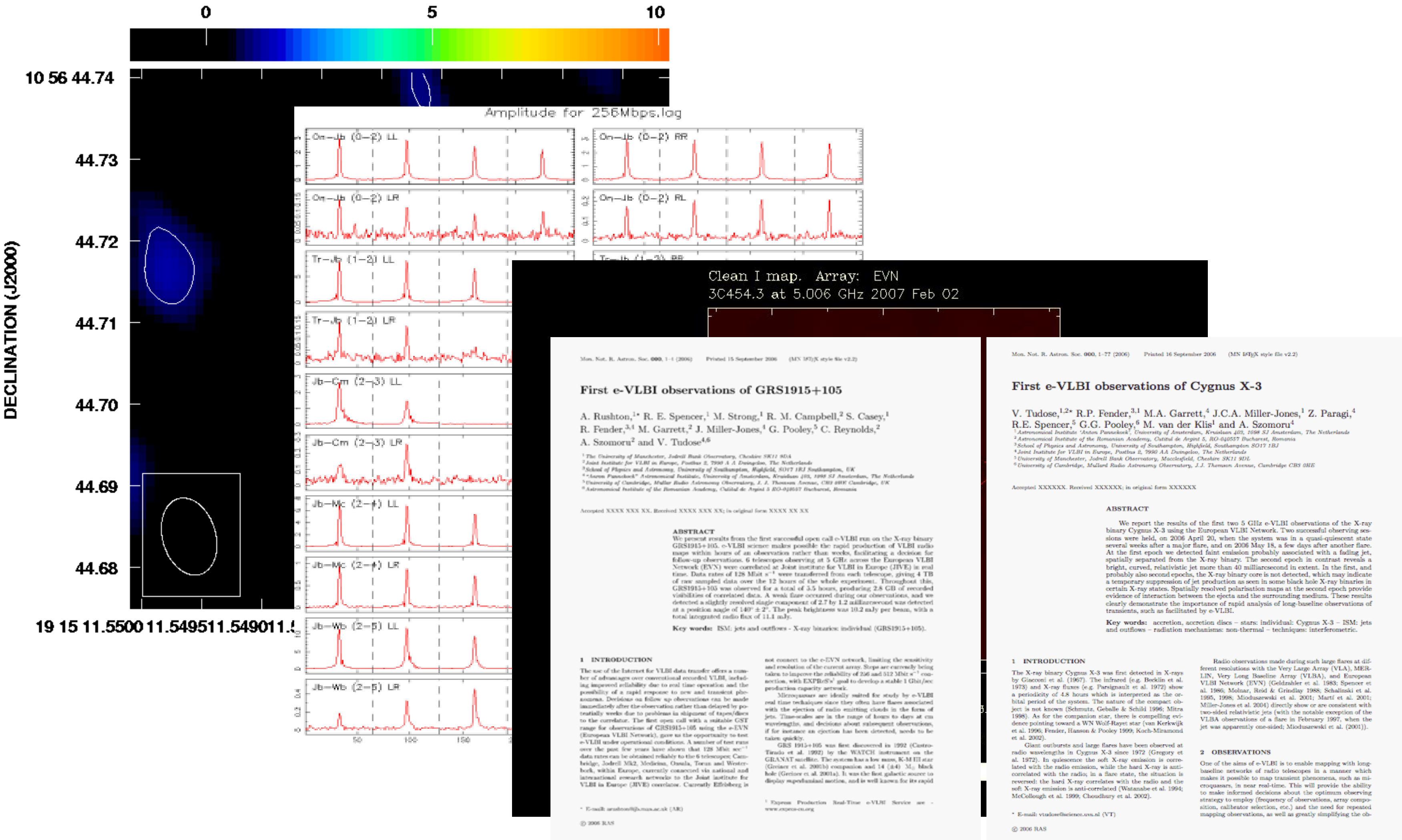
- Minimizing data loss by careful scheduling
- Increase of production data rate from 128 Mbps to 256 Mbps
- 5-station fringes at 512 mbps
- Inclusion of Metsähovi and Medicina telescopes



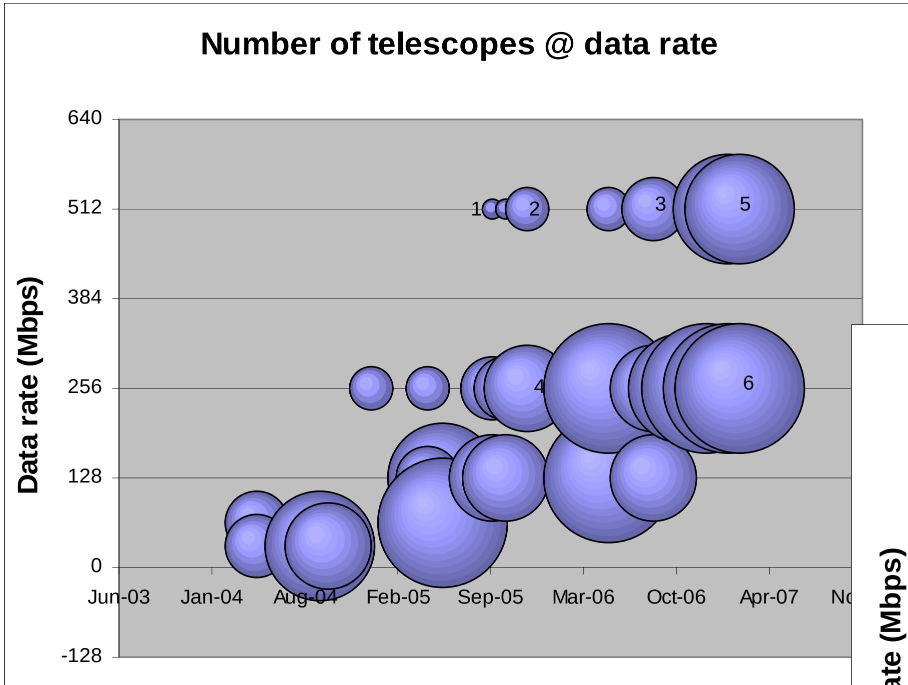
Test cases: black holes in outburst

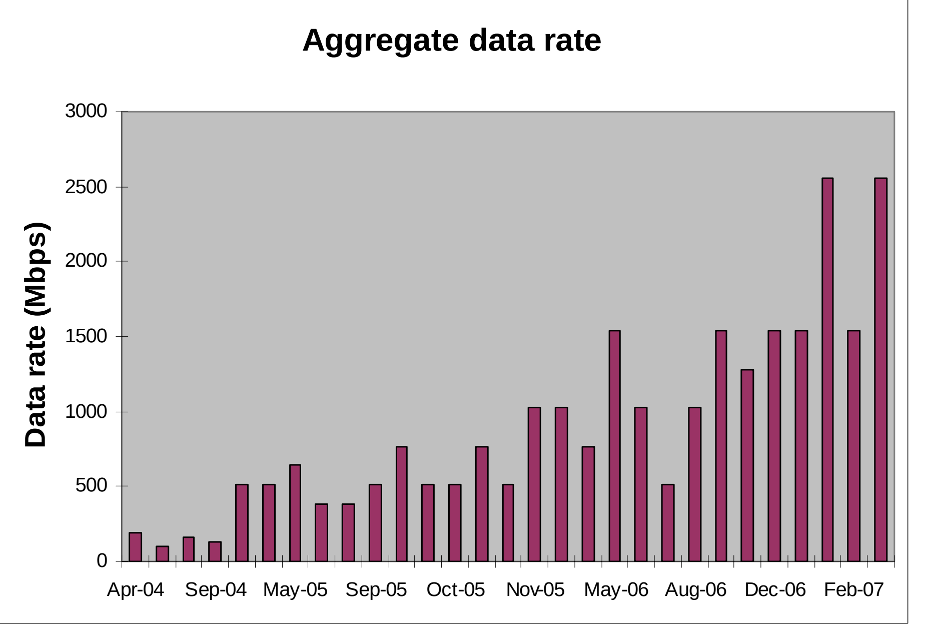
- Data processing took 1-2 weeks with first images <u>within 48 hours</u>
- Publication took less than 2 months

e-VLBI results



e-VLBI: data rate improvements

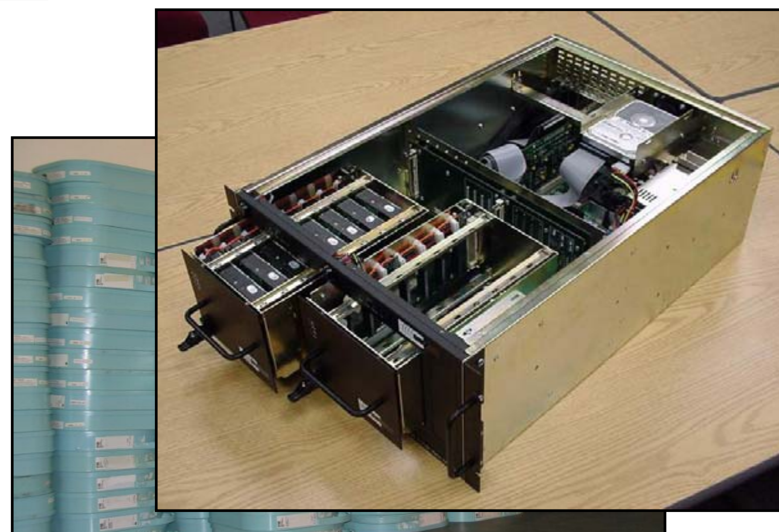




The EVN MarkIV correlator

Custom-made hardware, ~500000 lines of C++ code

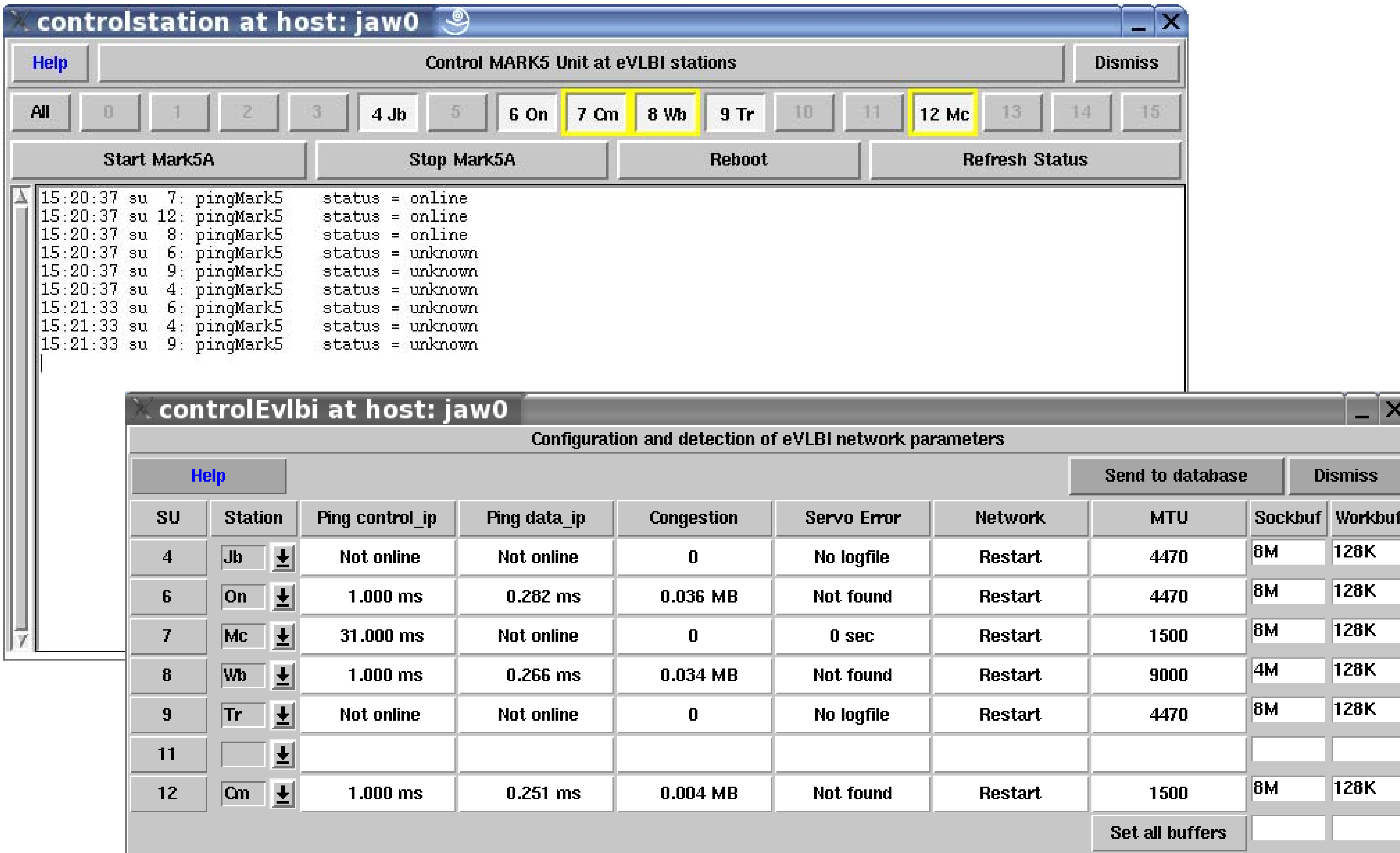
Designed and built on tape technology, only recently adapted for disk-based recording



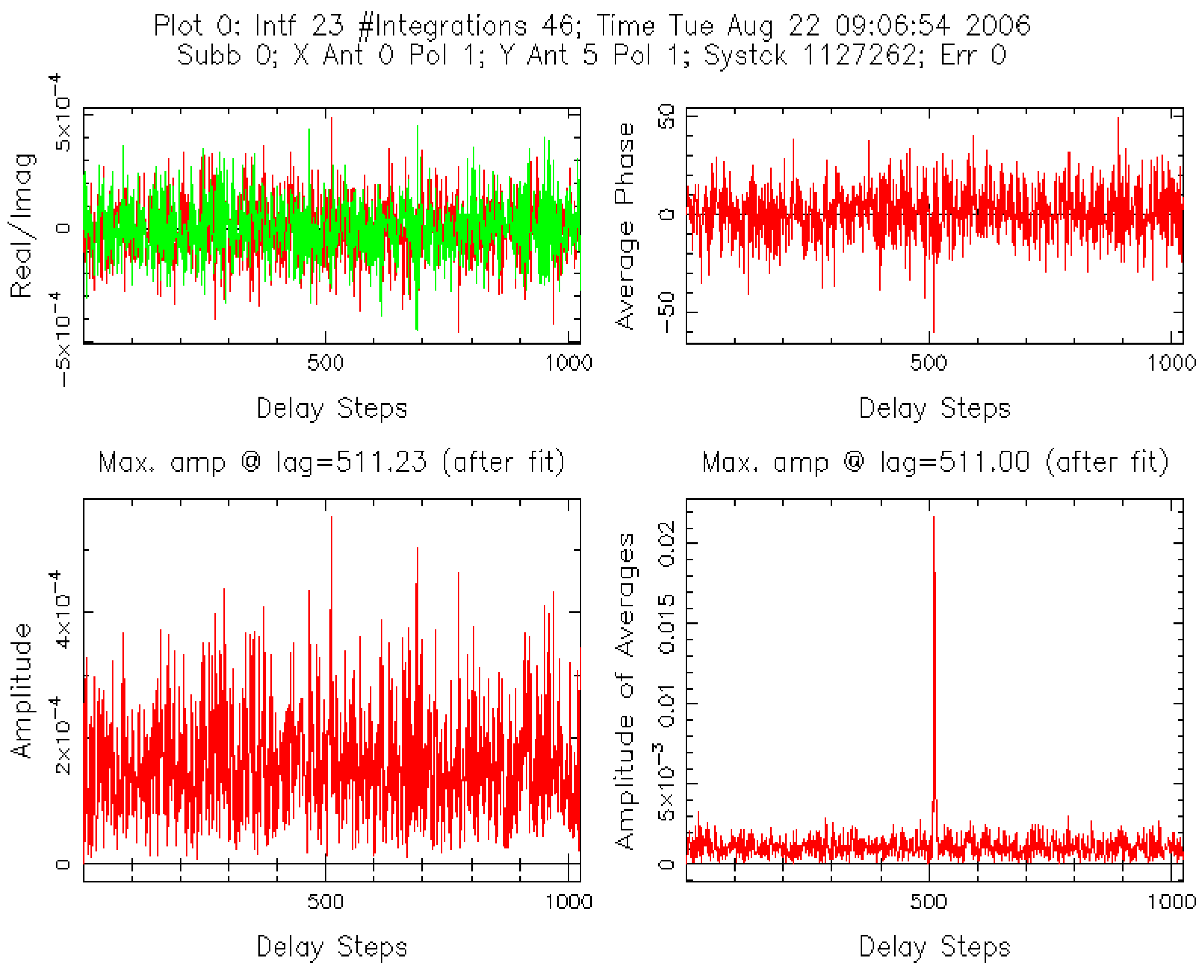
Work packages

- Standard e-VLBI correlator mode:
  - Wall-clock time vs ROT
  - Centralized control
- Rapid response/ToO functionality:
  - Real-time monitoring/feedback/analysis
  - On-the-fly adjustment of observational/correlator parameters
  - Improved amplitude calibration
- Data transport issues
  - Lightpaths: guaranteed (high) bandwidth
  - Transport protocols (modified TCP, UDP-based, VSI-e…)
- Hardware issues
  - New control computers (reliability, speed)
  - Replacement of SUs (cause of much misery)
  - Networking hardware (router, interfaces)

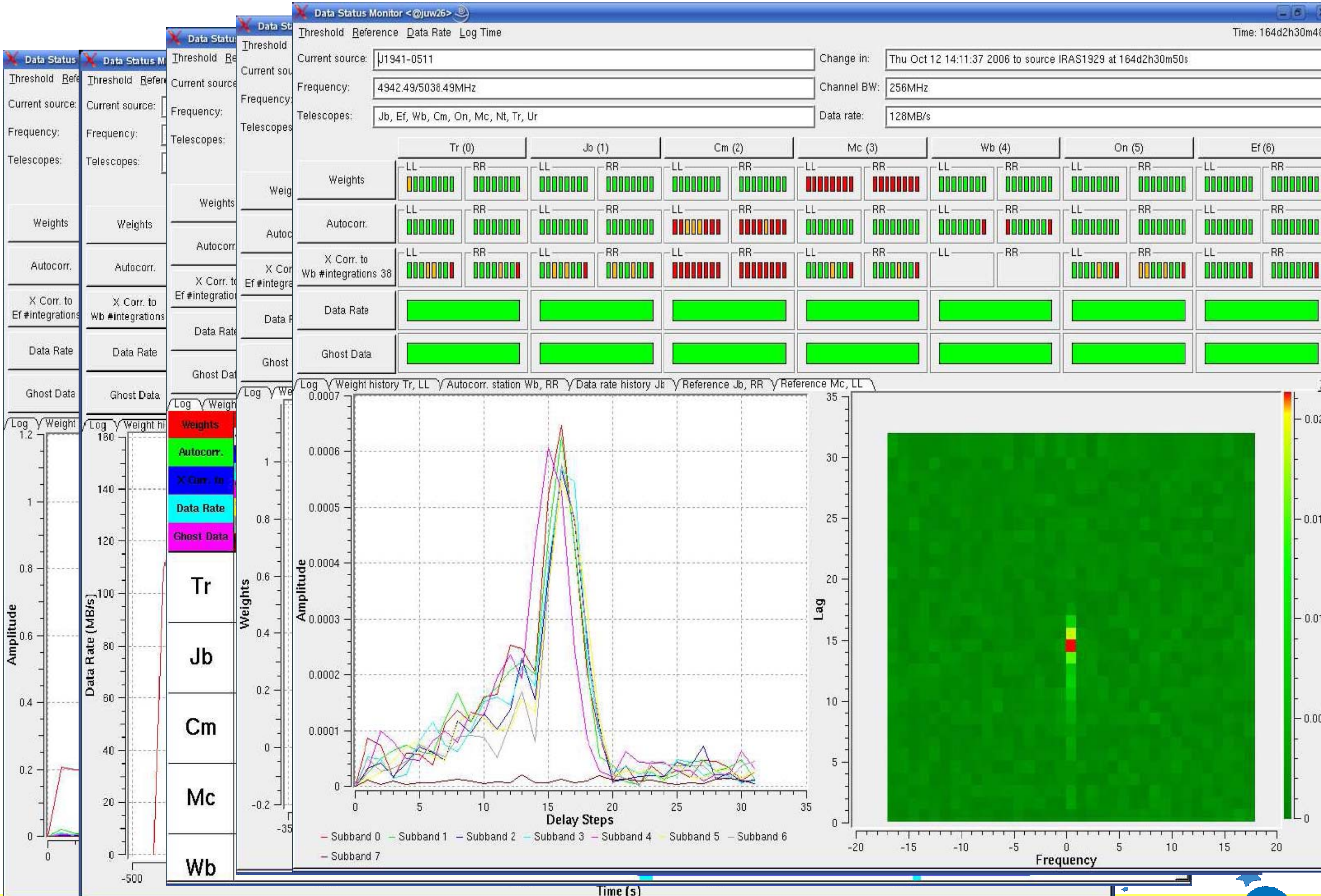
e-VLBI correlator mode: control interfaces



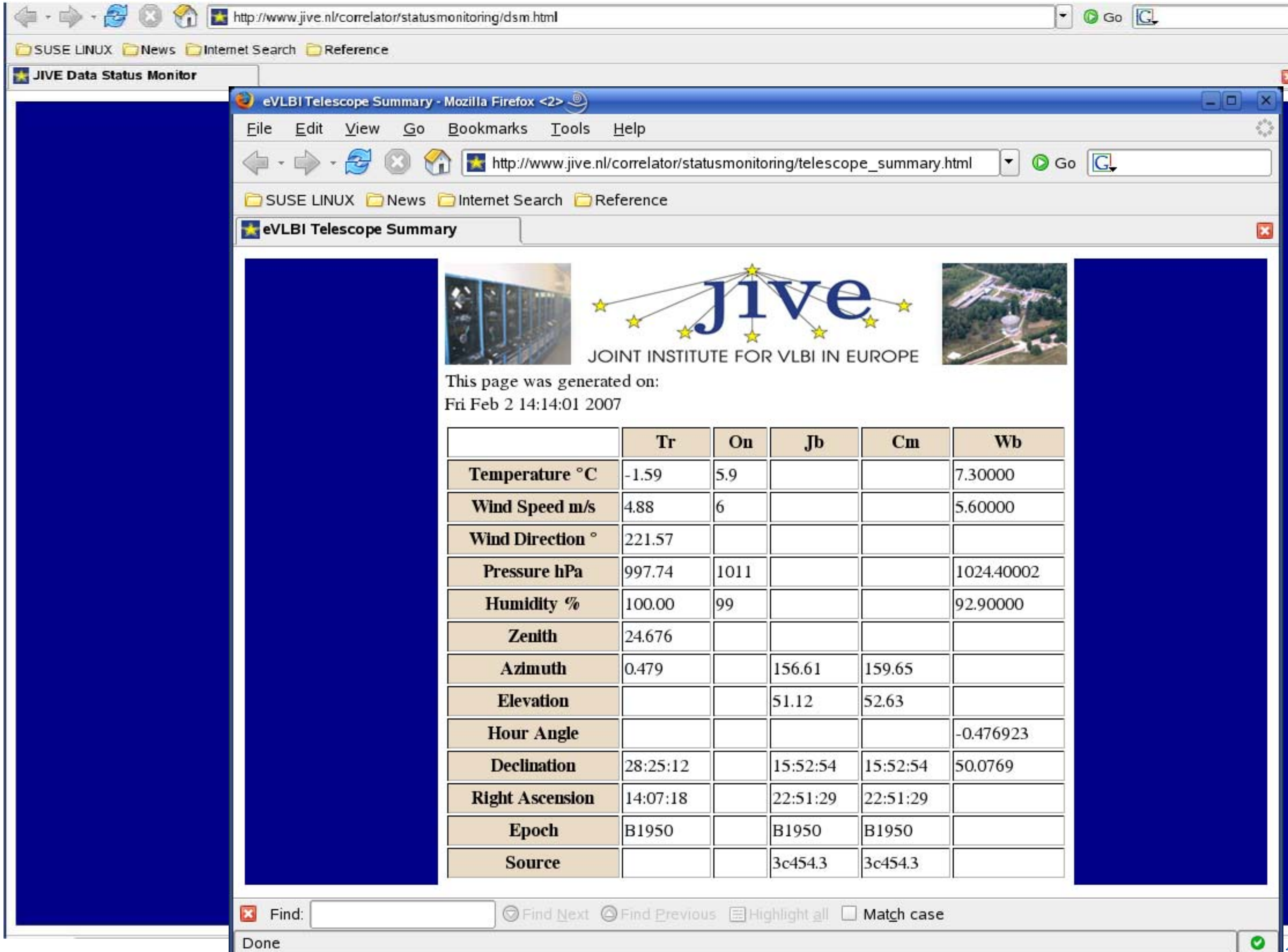
Monitoring tools: Integrating fringe display



Monitoring tools: Data status monitor



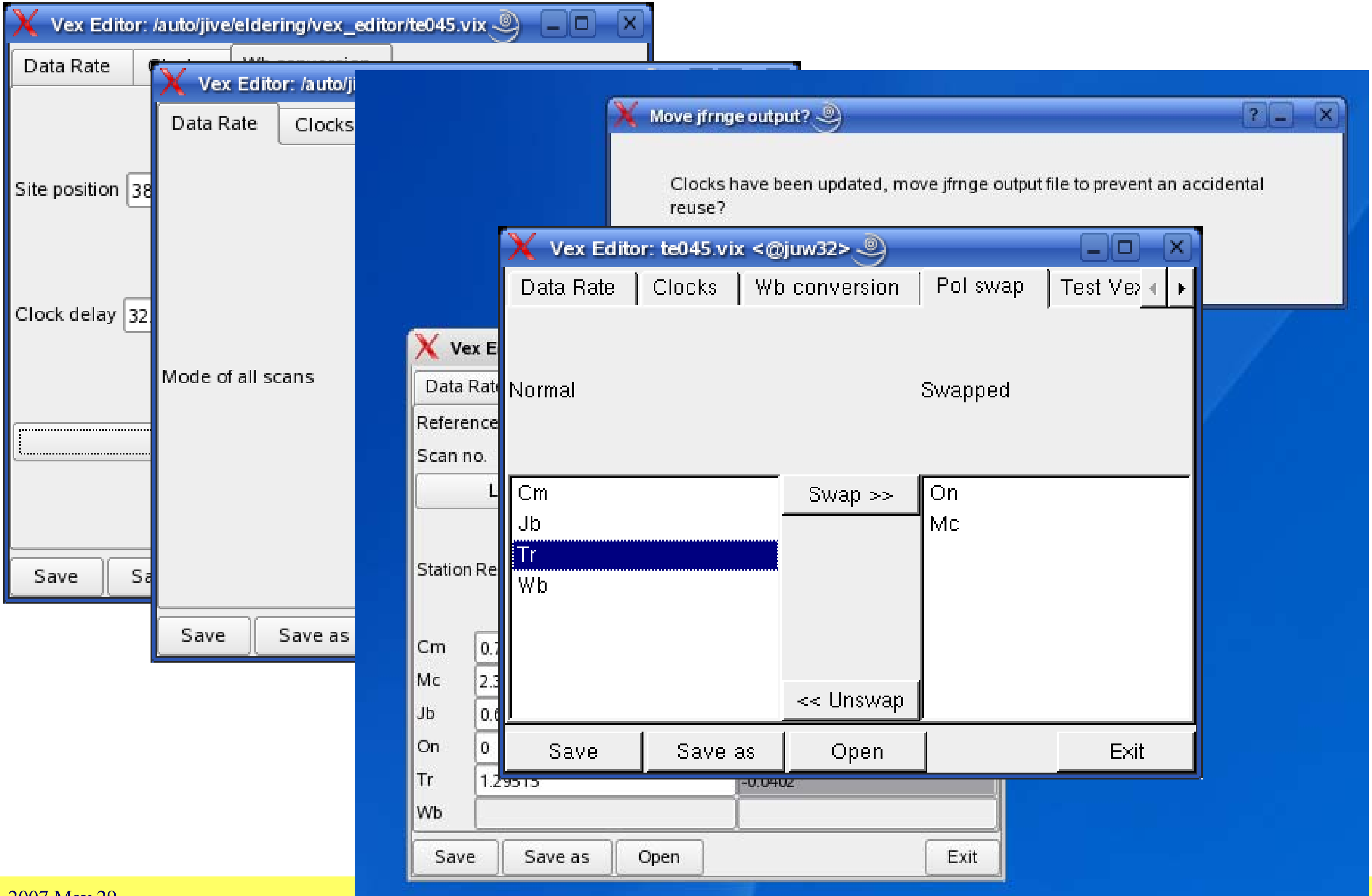
Monitoring tools: Data status monitor



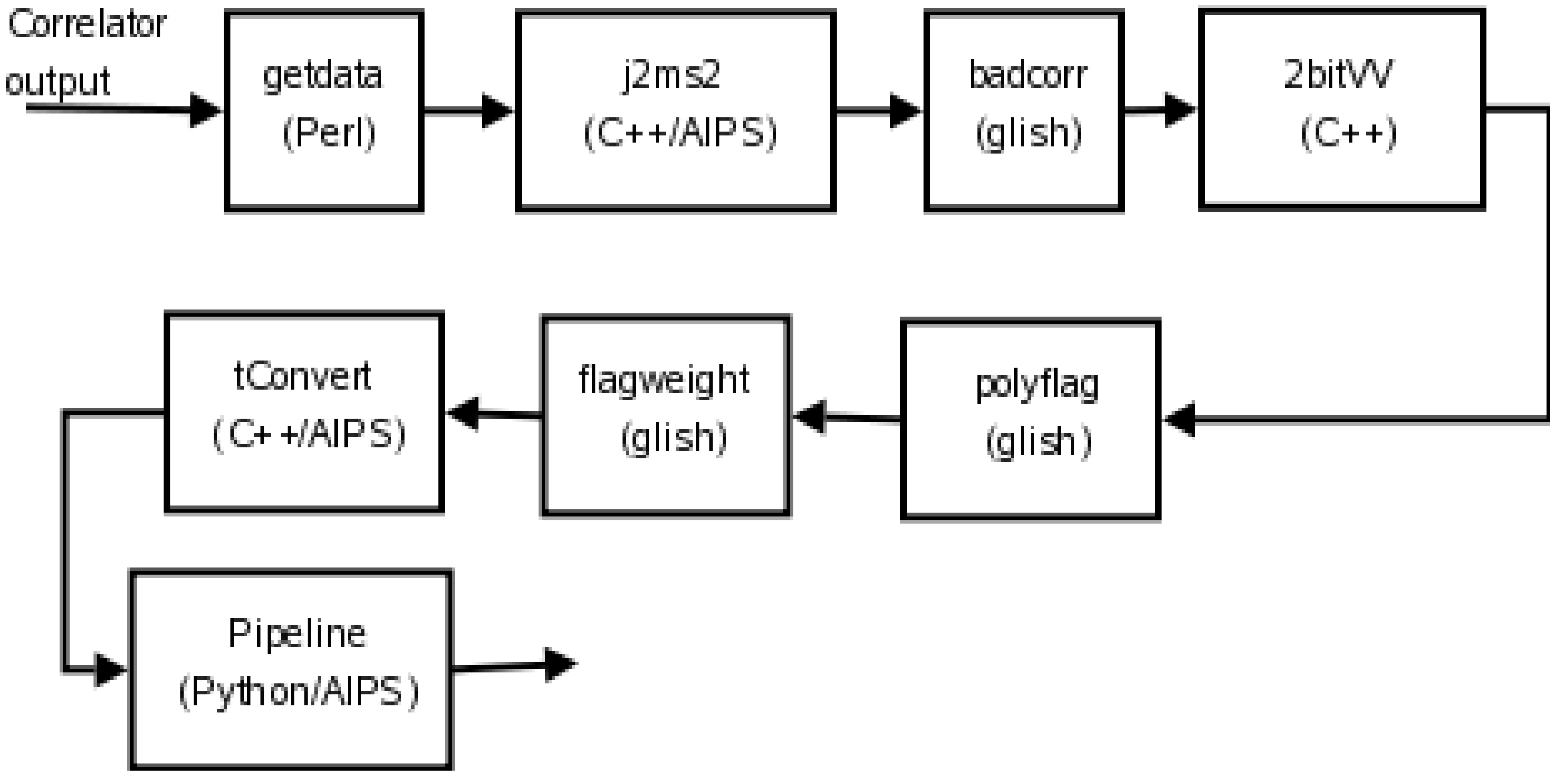
This page was generated on:
Fri Feb 2 14:14:01 2007

	Tr	On	Jb	Cm	Wb
Temperature °C	-1.59	5.9			7.30000
Wind Speed m/s	4.88	6			5.60000
Wind Direction °	221.57				
Pressure hPa	997.74	1011			1024.40002
Humidity %	100.00	99			92.90000
Zenith	24.676				
Azimuth	0.479		156.61	159.65	
Elevation			51.12	52.63	
Hour Angle					-0.476923
Declination	28:25:12		15:52:54	15:52:54	50.0769
Right Ascension	14:07:18		22:51:29	22:51:29	
Epoch	B1950		B1950	B1950	
Source			3c454.3	3c454.3	

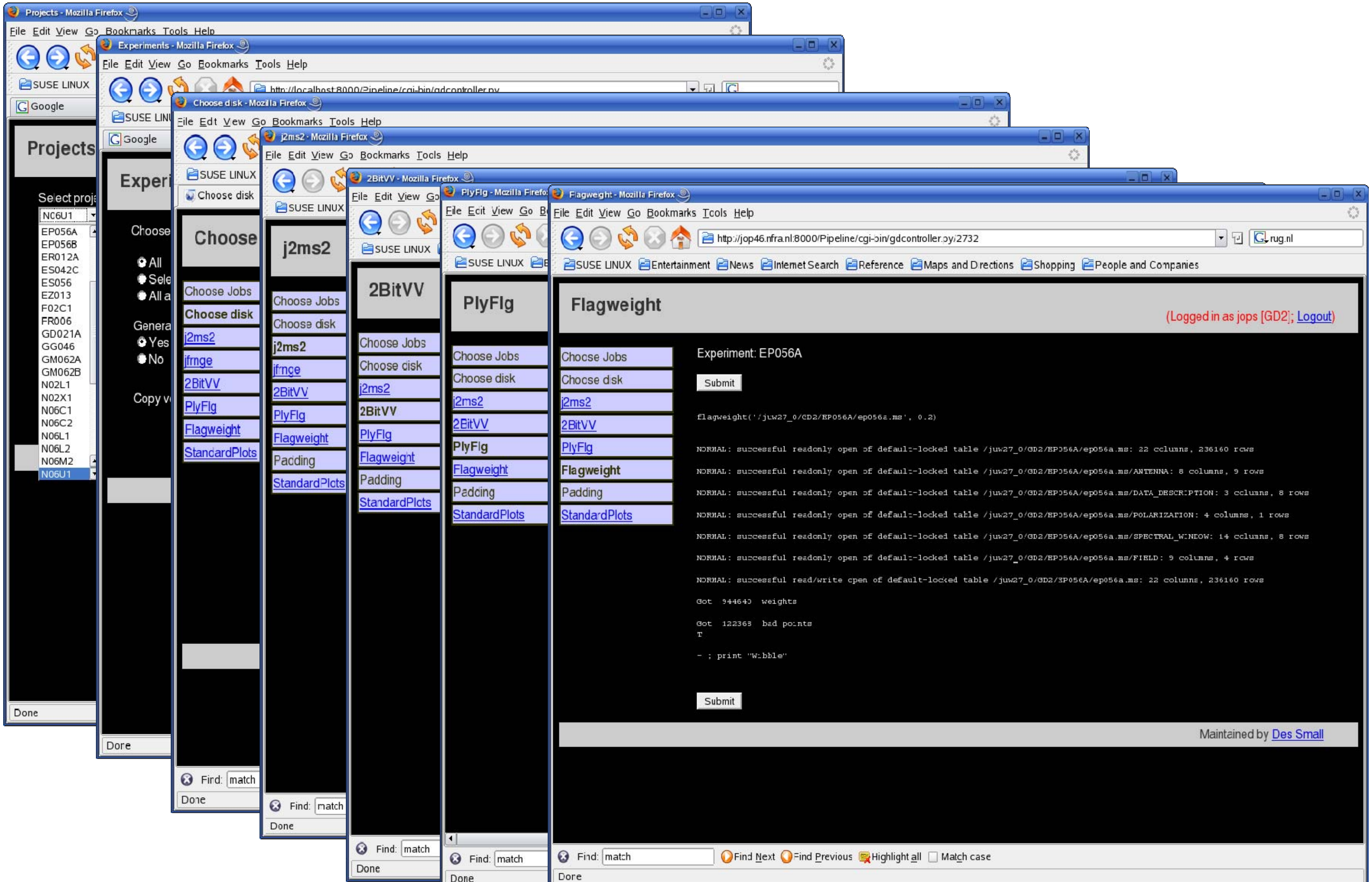
Rapid adjustment of correlator params: vex editor



(near) Real-time analysis: streamlining of post processing



Web-based Post-processing



Protocol work in Manchester:

- Protocol Investigation for eVLBI Data Transfer
- Protocols considered for investigation include:
  - TCP/IP
  - UDP/IP
  - DCCP/IP
  - VSI-E RTP/UDP/IP
  - Remote Direct Memory Access
  - TCP Offload Engines
- Work in progress – Links to ESLEA UK e-science
  - Vlbi-udp – UDP/IP stability & the effect of packet loss on correlations
  - Tcpdelay – TCP/IP and CBR data

Protocols (1)

Mix of High Speed and Westwood TCP (Sansa)

2.2.4 Our Proposed Algorithm (HTCP with Bandwidth Estimation)

Basing on the observations from the above stated algorithms [2, 4, 11, 6, 9, 8] as well as some comparative studies [1, 3], we propose an algorithm that combines the strengths of HTCP and TCPW. With this algorithm a connection will be established following HTCP-like rules, however at a packet loss, we propose that the TCPW's adaptive decrease be evoked. From our simulations with NS2 [7], on a packet loss event a TCPW flow sets its CWND to a value averagely 17% greater than that set by an HTCP flow. The increase factor will be controlled by the HTCP adaptive increase, which controls increment relative to the elapsed time since the last packet loss event. From our simulations the HTCP increase function yields much faster growth of the CWND than that obtained with TCPW. The aim is to use the tested bandwidth estimation mechanism employed by TCPW algorithm and combine it with adaptive increase mechanism of HTCP. The expectation is that this will ensure higher link utilisation than what is possible with any of TCPW or HTCP. The weaknesses of HTCP in short RTT flows as well as parallel flows revealed in [3] we propose to solve by employing the tested TCPW adaptive backoff .The non-adaptive nature of TCPW increase factor will be taken care of by the HTCP adaptive increase factor. Our proposed algorithm is summarised in (17) and (19).
On Acknowledgement:

$$CWND \leftarrow CWND + \alpha/CWND \qquad (17)$$

Where

$$\alpha_i \leftarrow \begin{cases} 1 & \triangle_i \leq \triangle^L \\ 1 + 10(\triangle_i - \triangle^L) + (\frac{\triangle_i - \triangle^L}{2})^2 & \triangle_i > \triangle^L \end{cases} \qquad (18)$$

Where $\triangle_i$ is the elapsed time since the last congestion event experienced by the flow i, $\triangle^L$ is the time duration used as the threshold for switching from the low to high speed regimes.

On Packet Loss event:

$$CWND \leftarrow B \times RTT_{avg} \qquad (19)$$

Where

$$B = CWND/RTT_{avg} \qquad (20)$$

Protocols (2)

Circuit TCP (Mudambi, Zheng and Veeraraghavan)

Meant for dedicated end-to-end circuits, fixed congestion window

No slow start, no backoff: finally, a TCP rude enough for e-VLBI?

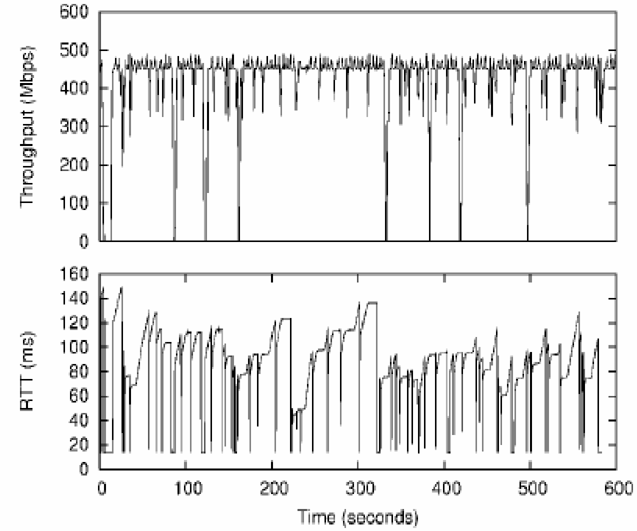


Figure 4. Throughput and RTT using BIC-TCP

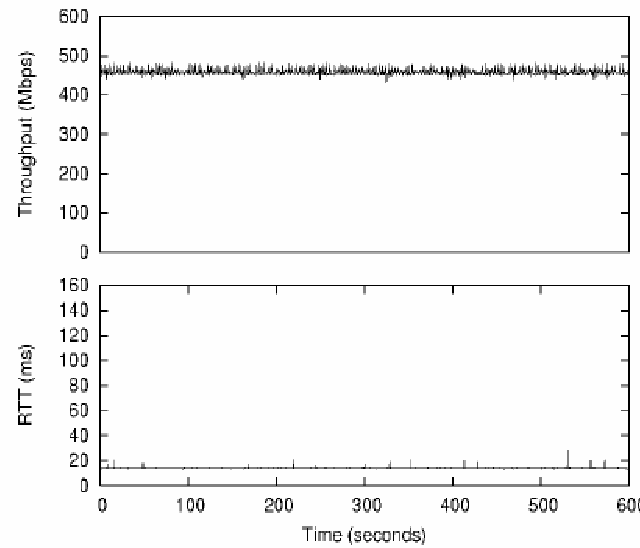


Figure 5. Throughput and RTT using C-TCP

Protocols (3)

Home-grown version of CTCP using pluggable TCP congestion avoidance algorithms in newer Linux kernels (Mark Kettenis)

Rock-steady 780 Mbps transfer using iperf from Mc to JIVE

Problem with new version of Mk5A software under newer kernels; should be solved by the end of May..

Hardware

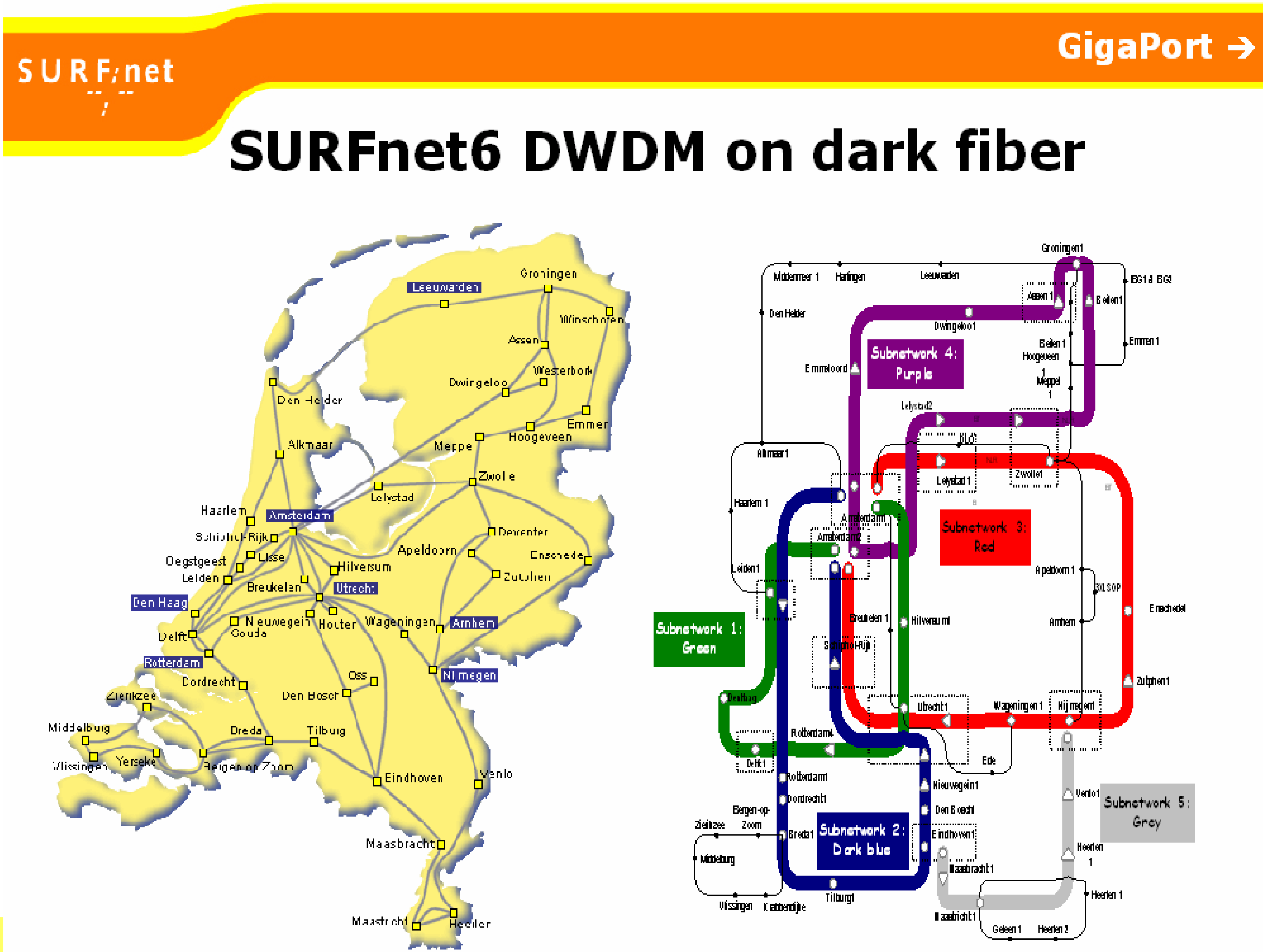
New control computers

- Solaris AMD servers, redundant power supplies
- Interchangeable for maximum reliability
- Cut down dramatically on (re-)start time
- Removal of single points of failure

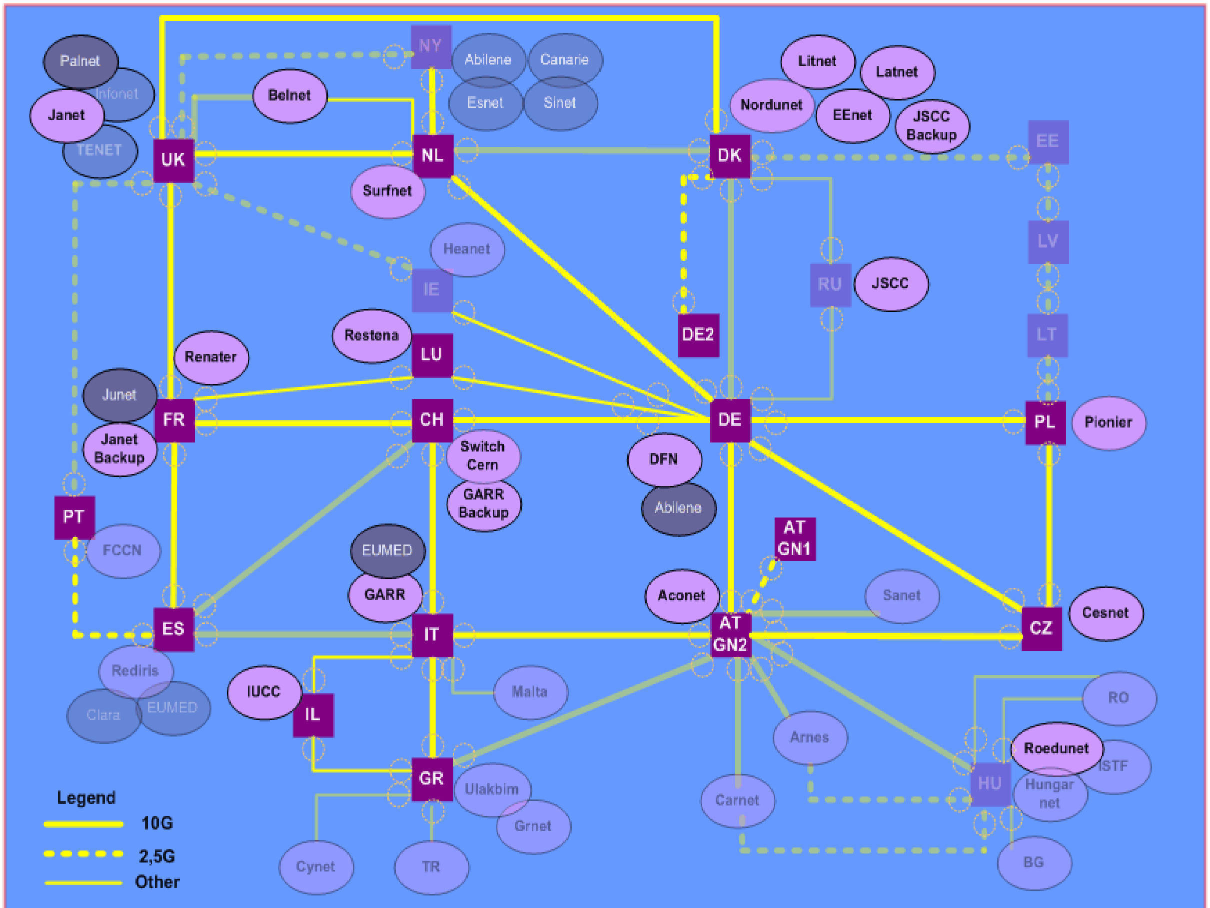
Mark5 upgrade

- Original Mark5s woefully under-powered (Pentium PIII…) New dual processor motherboards, memory, power supplies
- Immediate result: Mc and Tr now sustain 512 Mbps
- Mark5A→B: new streamstor cards, new serial links (optical and coax), Correlator Interface Boards
- Elimination of Station Units, VSI compliance

Hybrid networks in the Netherlands..



..and across Europe: GÉANT2 network upgrade



EXPReS

Local solutions: Tr connectivity bottleneck (partially) solved

Black Diamond 6808 switches:

New interfaces (10GE) system in old architecture (1GE)

- Originally 8x1GE interface per card
- 10GE NIC served by 8 x 1GE queues
- Queuing regime – RR (packet based) and flow-based

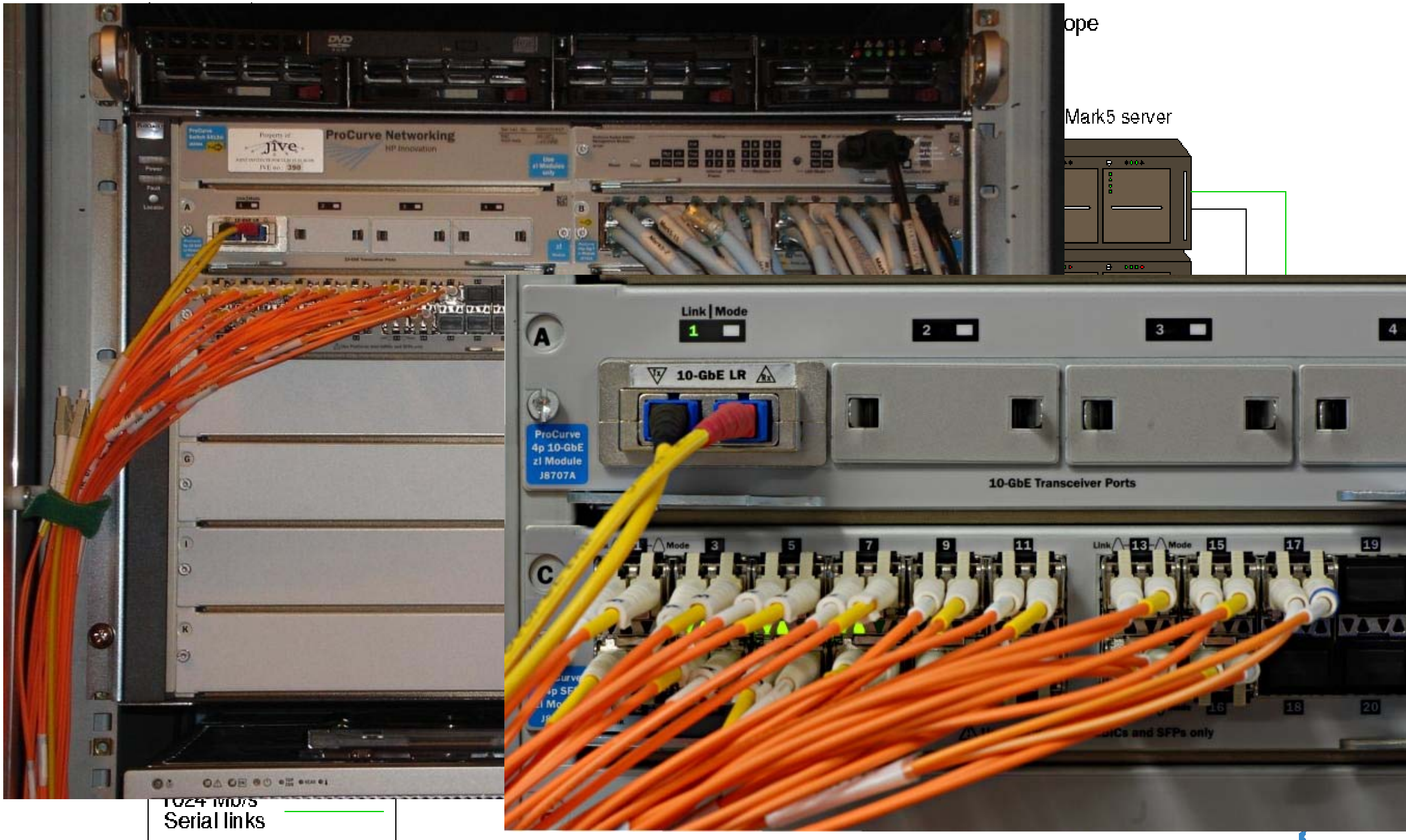
Flow based:

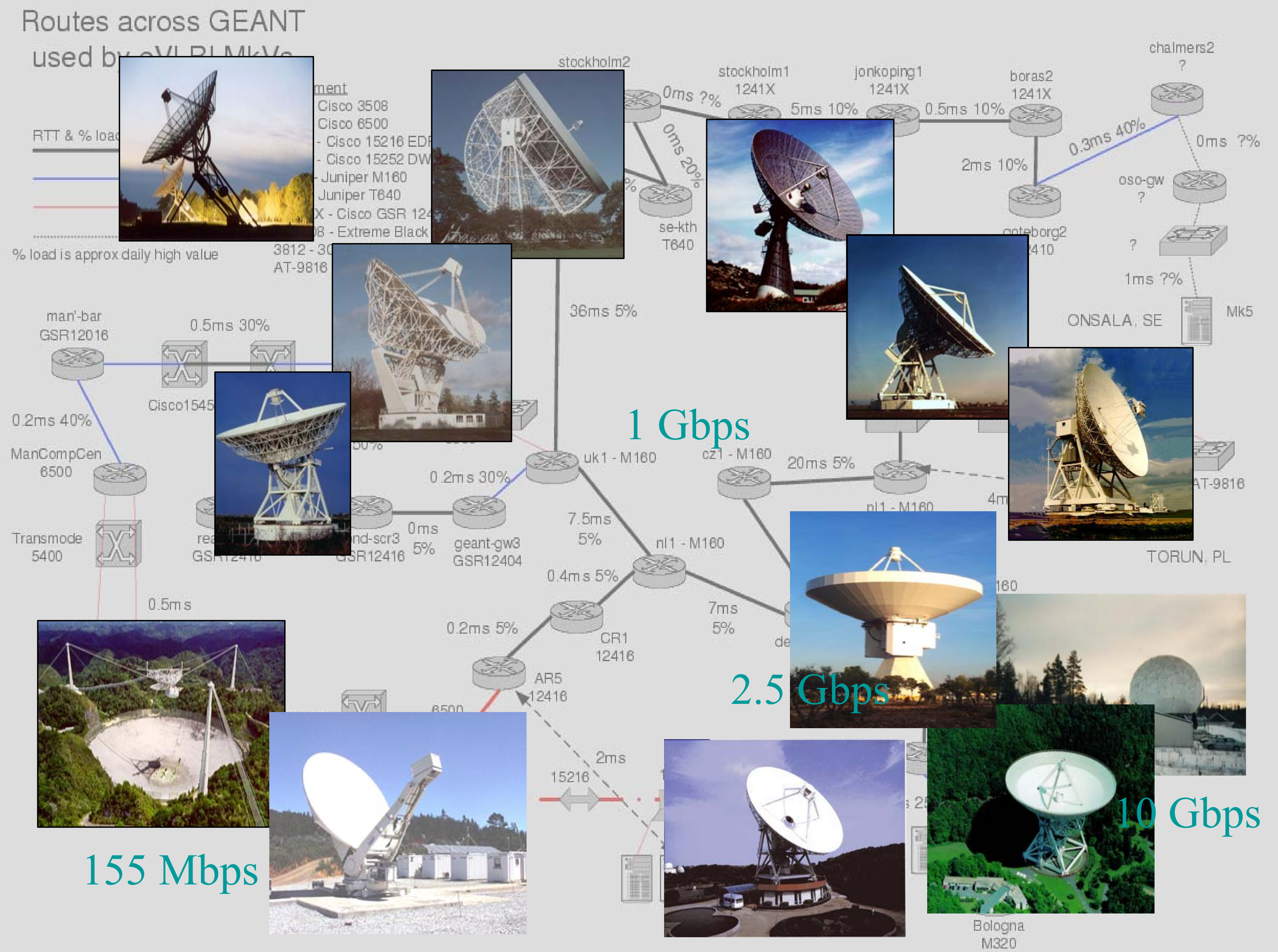
Max. flow capacity – 1Gbit/s – background traffic.

There is no known reordering workaround to solve this problem.

But, re-routing of traffic between Poznan and Gdansk has made a big difference.

Upgrade of network at JIVE





Routes across GEANT
used by
RTT & % load
% load is approx daily high value
Cisco 3508
Cisco 6500
- Cisco 15216 EDF
- Cisco 15252 DW
- Juniper M160
Juniper T640
X - Cisco GSR 124
8 - Extreme Black
3812 - 3
AT-9816
stockholm2
0ms ?%
stockholm1
1241X
5ms 10%
jonkoping1
1241X
0.5ms 10%
boras2
1241X
chalmers2
?
0.3ms 40%
0ms ?%
0ms 20%
2ms 10%
se-kth
T640
oso-gw
?
goteborg2
?
1ms ?%
ONSALA, SE
Mk5
man'-bar
GSR12016
0.5ms 30%
Cisco1545
36ms 5%
1 Gbps
0.2ms 40%
ManCompCen
6500
uk1 - M160
cz1 - M160
20ms 5%
AT-9816
0.2ms 30%
Transmode
5400
0ms
5%
geant-gw3
GSR12404
GSR12416
7.5ms
5%
nl1 - M160
TORUN, PL
0.4ms 5%
0.5ms
0.2ms 5%
CR1
12416
7ms
5%
AR5
12416
6500
2.5 Gbps
2ms
15216
10 Gbps
155 Mbps
Bologna
M320

SA1: the next 18 months

All personnel in place, at JIVE and JBO

Main focus:

- Adaptive scheduling
- Increasing the operational data rate
- Improving flexibility and robustness of correlation
- New telescopes

- Inclusion of eMERLIN telescopes

- Re-assessment/refinement of deliverables

Specific tasks

Crucial:

- fast/adaptive (re-)scheduling
  - Transform EVN into a truly flexible instrument
  - Reaction time of hours
  - Remote control of widely different telescope control systems
  - Safeguard local operational constraints

- hard/software upgrade
  - Phase out aging Station Units, main cause of operational instability
  - Involves large software effort
  - VSI compliance

Fine-tuning of deliverables

- Use of WSRT synthesis data for e-VLBI calibration (absolute flux calibration, polarization calibration, source selection)
- Space craft tracking correlator mode
- On-the-fly fringe fitting
- Real-time download and extraction of station information
- Automated correlator diagnostics
- Removing/adding stations from the correlation process on the fly
- Investigating a 1024M sub-array

Connectivity improvements:

- Use of new/modified transport protocols, lightpaths across Europe
- Increase granularity of data streams to optimize utilization of available bandwidth
- Increase of sensitivity and resolution through addition of telescopes: Effelsberg, Yebes, Shanghai

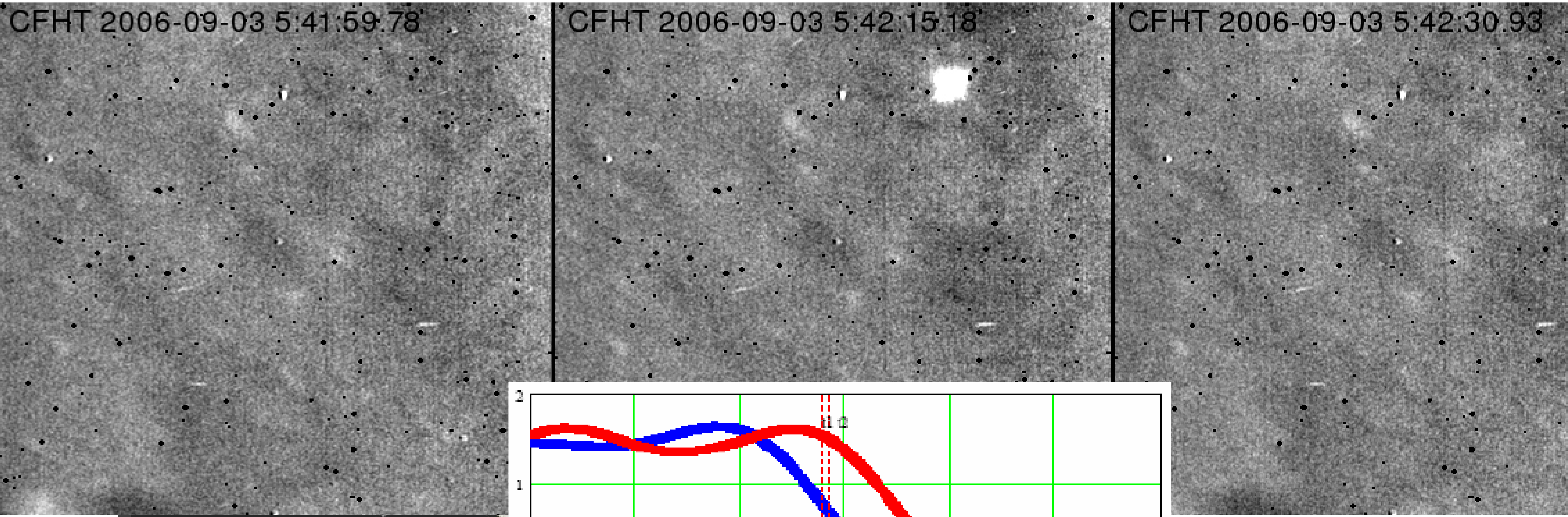
- Improving global connectivity: South-America, Puerto Rico, China, Australia

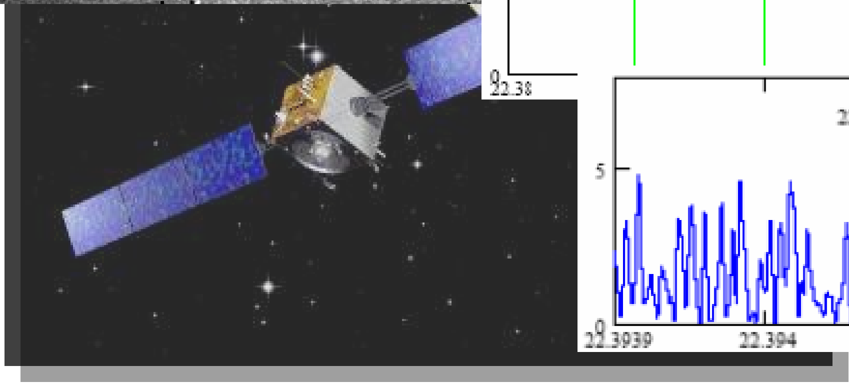
Global connectivity:

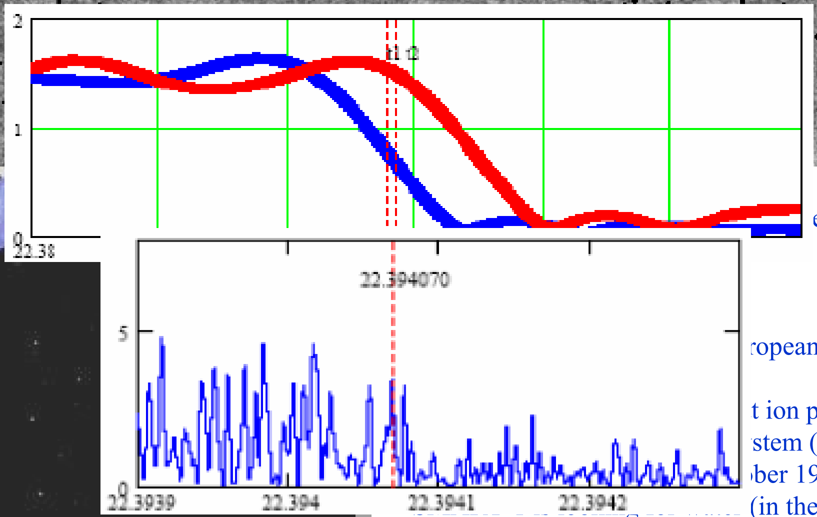
- Arecibo: 64 Mbps real-time operations in the past, currently < 32 Mbps. New submarine cable facilities in near future…
- Australia: 4 LBA telescopes connected at 1 Gbps. AARNET currently investigating transport to Europe. Issues with format conversion, command interfaces. First test planned before summer 2007
- South Africa: local situation improving, connection to Europe remains problematic

e-VLBI to South America: SMART-1

SMART-1 factsheet







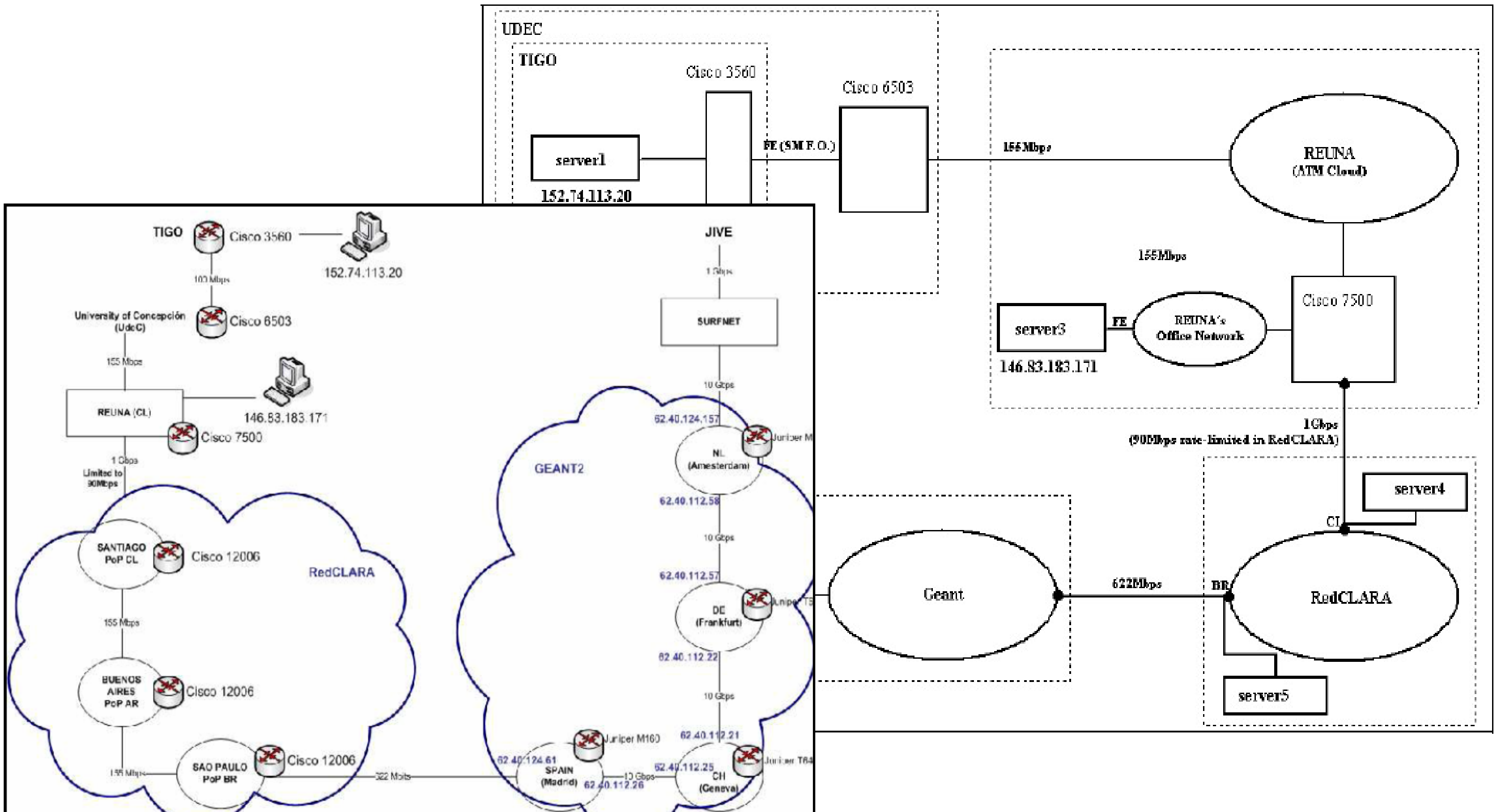
ember 2004. Conducting lunar

ropean spacecraft to travel to and

t ion propulsion has been used as a
stem (the first was NASA's Deep
ber 1998).

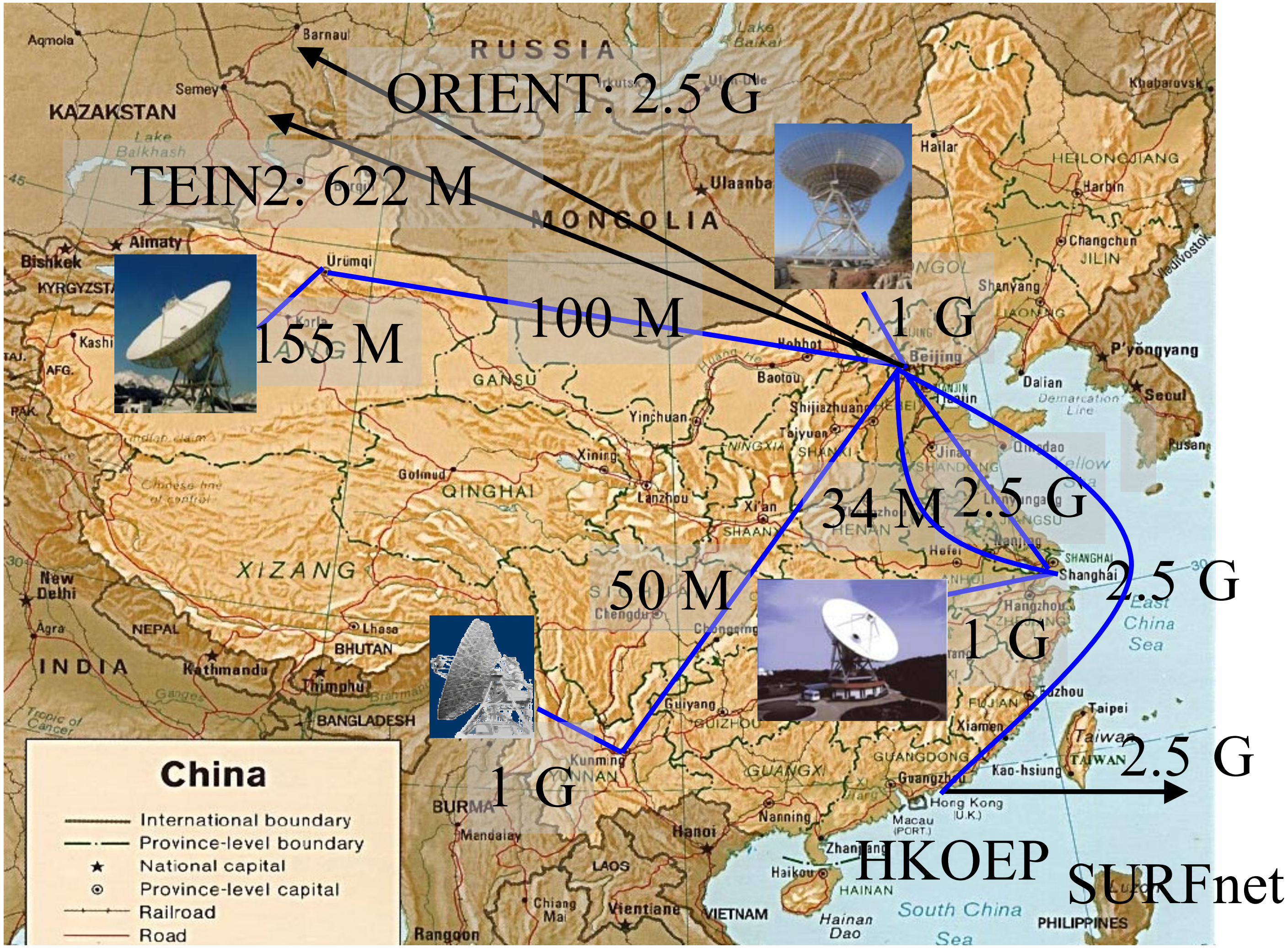
(in the form of ice) on the Moon.
To save precious xenon fuel, SMART-1 uses 'celestial mechanics', that is, techniques such as making use of 'lunar resonances' and fly-bys.

e-VLBI to South America (2)

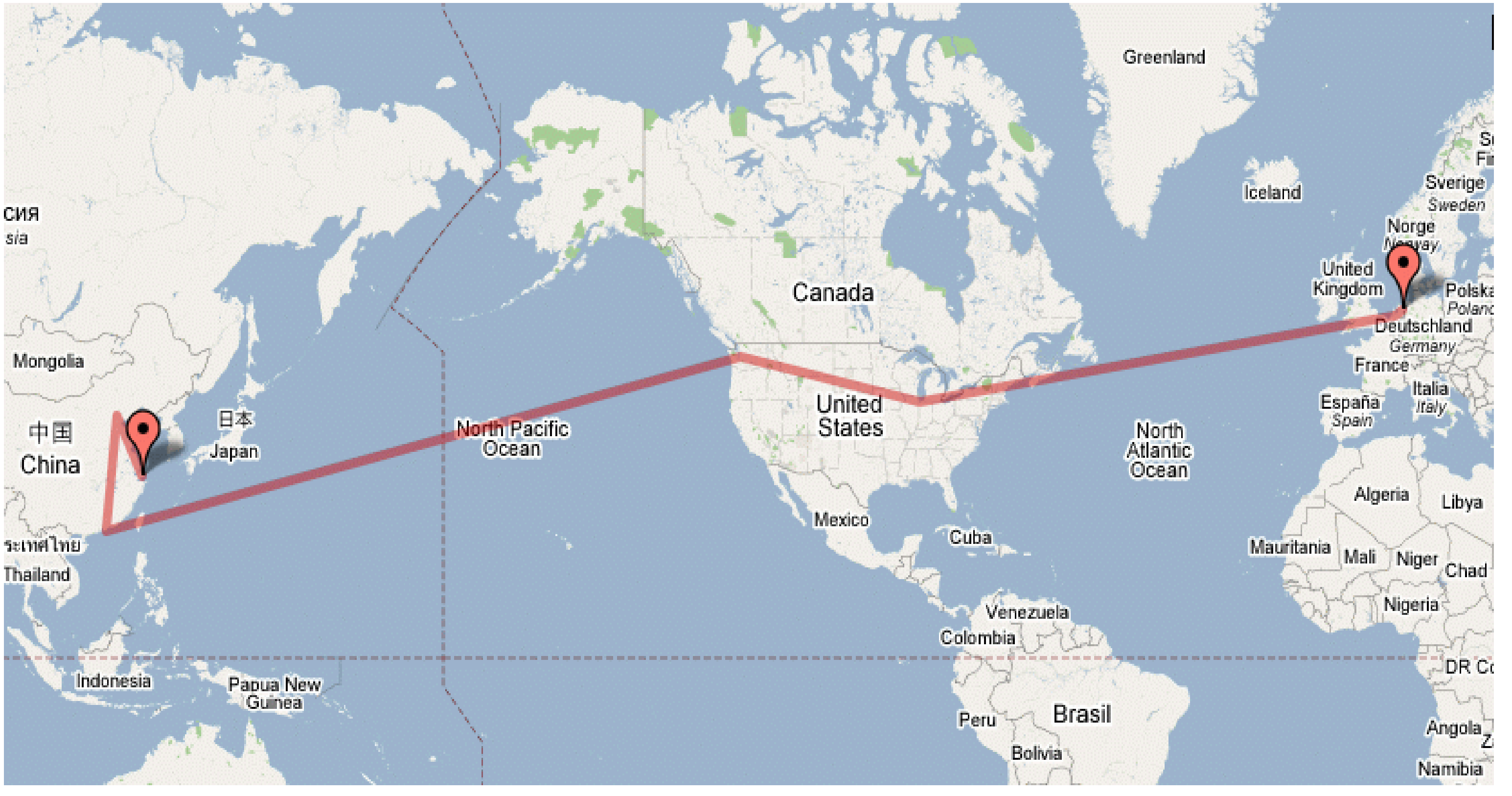


Server3 to server2: 274ms
Server4 to server2: 273ms
Server5 to server2: 226ms

e-VLBI to China

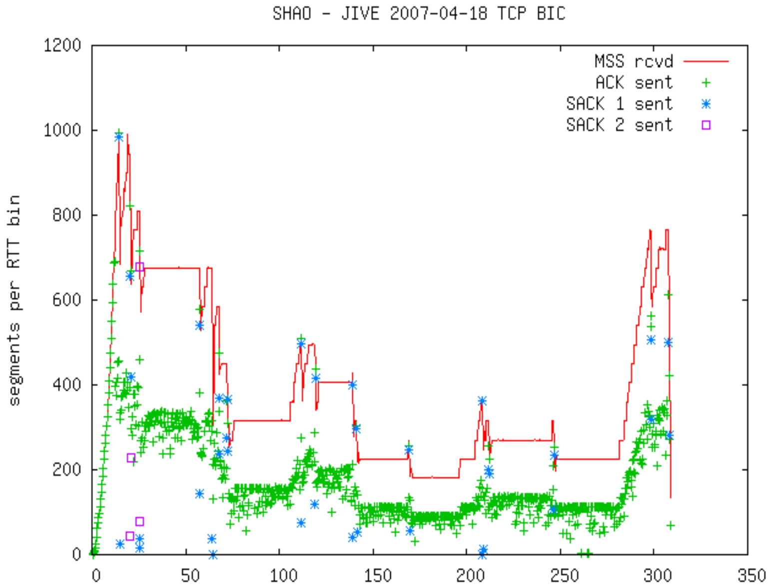


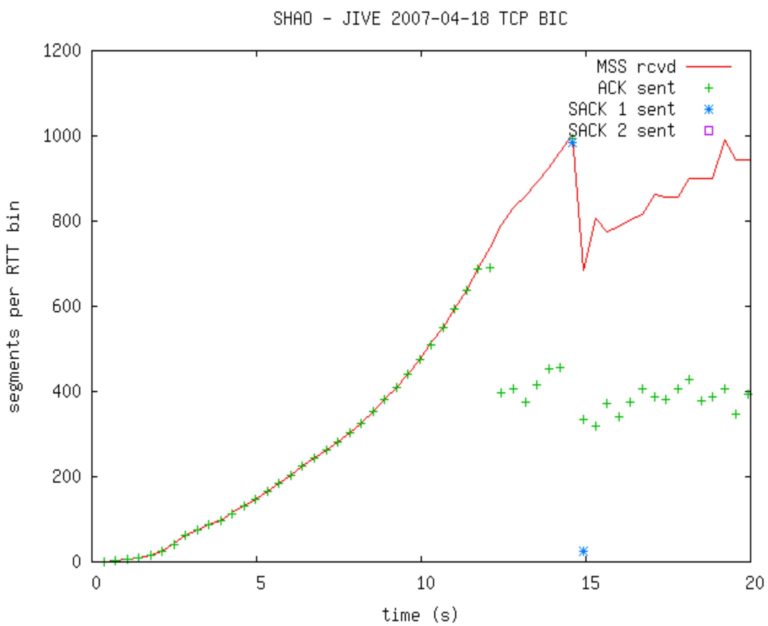
e-VLBI to China (2)

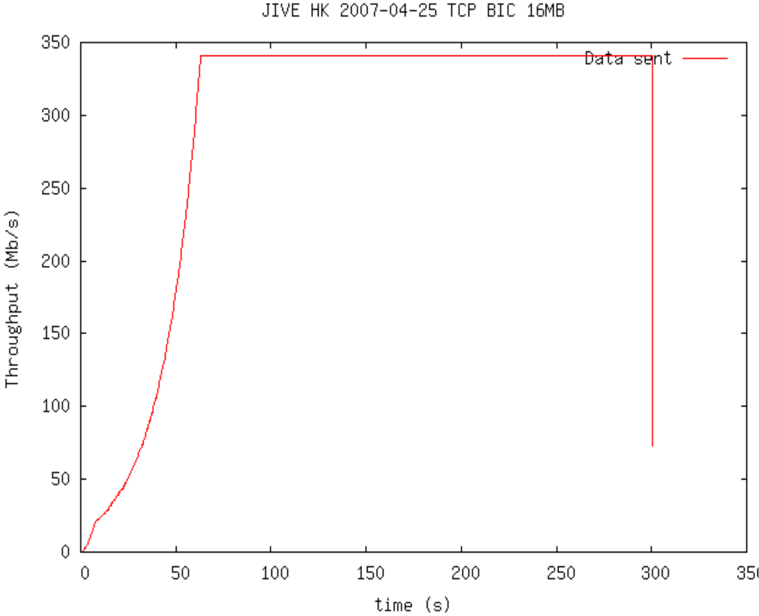


Connection Sheshan (near Shanghai) – JIVE (via HongKong, Chicago).
RTT 360 ms, real distance 8800 km, path length 27500 km

e-VLBI to China (3): Connectivity tests Sheshan-JIVE







e-VLBI to China (4)

- First tests (iperf with UDP, to rule out tuning issues):
  - HongKong – JIVE: 577 Mbps (lightpath 622 Mbps)
  - Sheshan – Beijing: 560 Mbps (822 Mbps with TCP!)
  - Beijing – HongKong: 1.85 Mbps (!)
- Follow-up
  - HongKong – JIVE: steady 320 Mbps with TCP
  - HongKong – Beijing: most lossy part of connection
- Latest result: nearly 800 Mbps SHAO-Hong Kong with TCP