JIVE Network design

Paul Boven

March 30, 2007

1 Introduction

Within the EXPReS project, the network at JIVE will be upgraded to support e-VBLI with more radio telescopes and at higher speeds. This document describes the reasoning behind the structure of the new JIVE network. It includes the IP plan, security considerations, IP subnetting and hardware decisions.

1.1 Current situation

At the moment, JIVE does not own or control any networking equipment. Physically, our Mark5's are connected by LX fiber to a SURFnet Nortel OMS6500 in the Astron/JIVE building, giving us 7x 1Gb/s fibers to the Netherlight equipment (Cisco 6500 switch/router) in Amsterdam. This limits our current e-VLBI capability to 7x 1Gb/s. In addition to that, we have a 1Gb/s dark fiber connection to the Westerbork Synthesis Radio Telescope (WSRT), also for e-VLBI.

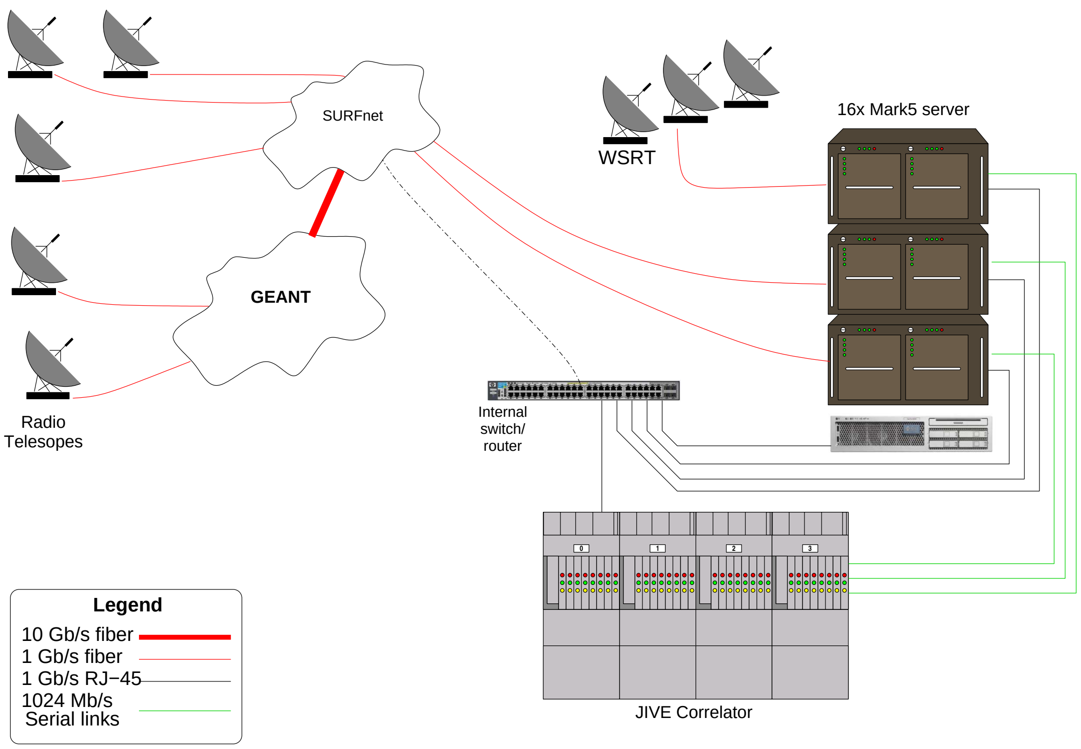

Figure 1: JIVE network before upgrade

2 New network configuration

SURFnet will upgrade our network connectivity in the months to come. They will supply us with a 10Gb/s link for IP routed traffic (rate-limited to 5Gb/s) and 16 lightpath connections of 1Gb/s. All the 1 Gb/s connections to SURFnet will be made using short-range multi-mode (SX) fiber interfaces, the 10Gb/s unfortunately needs to be a single-mode fiber. All this traffic needs to be delivered to our 16 Mark5 stations - cheapest would be to use Cat5-cabling to those. In between will be our new switch/router that needs to be able to switch and route these traffic flows.

2.1 Local network

The Allied Telesys 24-ports 100Mb/s switch in the correlator rack is full. As part of our networking upgrade, we would want to replace this switch with a faster one, with more ports. Initially we were considering to integrate this functionality with the new central switch/router. But because these are networks with completely separate address-spaces and routing setups, and as none of the candidate switch/routers currently supports VRF (Virtual Router Functionality) it turned out to be neccesary to use a separate switch/router for the local network. Even so, upgrading the current switch with a new switch/router simplifies our local network and allows us to eliminate some bottlenecks, such as the aging Jbrouter PC-based router from the correlator room, and take over the routing functionality for the PCint. Replacing these two routers and integrating them in the new switch also means that we will no longer need to go trough the NFRA network and back in order to run the correlator.

2.2 Bandwidth requirements

The JIVE correlator can handle up to 16 stations simultaneously, at speeds of up to 1024Mb/s. The sustained forwarding rate of the new switch should therefore be well above 16Gb/s. VLBI data currently comes in powers of two, e.g. 256Mb/s, 512Mb/s and 1024Mb/s. We can already achieve 512Mb/s with several of the stations, but 1024Mb/s unfortunately doesn't fit trough a single 1Gb/s ethernet connection. Trunking two ports together for 2Gb/s bandwidth might seem a sensible solution at first, but turns out not to work for e-VLBI data: Because this data is sent inside a single TCP connection (flow), an 802.3ad (LACP) trunk is not able to distribute that traffic over both trunk members. Most other proprietary trunking methods offered by the switch manufacturers suffer from the same limitations. This is a design limitation in most trunking methods in order to prevent the packets from one TCP-session overtaking each other, and arrive out of order at the destination host.

Network interfaces with 10Gb/s are becoming readily available and their price has started to come down to about €700. Connecting at least some of the Mark5 at this speed has become a viable option to enable 1024Mb/s e-VLBI operations. This would require a 10Gb/s PCI-E interface card for those Mark5s, and a matching switch port on the networking equipment. At this time, 10Gb/s ethernet is available on fiber, or on CX-4 (copper, infiniband). CX-4 is a bit cheaper than using fiber but requires some rather cumbersome special cables. But the new 10Gbase-T standard (802.3an) has just been ratified, and the first 10Gb/s RJ-45 network cards are becoming available. This promises to become the most cost-effective way of connecting Mark5's to the switch at 10Gb/s. The PCI-E networking cards are already available, and 10Gbase-T interfaces for switches are expected before the end of the year. We have therefore opted to do all the new copper wiring using Cat-6a cabling, which is rated for 10Gb/s throughput, and is compatible with 1Gb/s as well.

2.3 Westerbork

The Mark5 at the WSRT has a 1Gb/s interface connected to a Cisco switch. This switch is used to drive an ZX (long-range single-mode, 70km) SFP. In Dwingeloo is a similar switch, converting the signal back to SX (short range, multi-mode) fiber which is then connected to one of the Mark5 that has an SX interface (all the other eVLBI capable Mark5's currently have LX interfaces). The dark fiber between the WSRT and Dwingeloo is 34.4km long and has a worst-case loss of 8.3 dB at 1500nm.

The easiest way to connect the WSRT to our new switch would be to plug the current SX fiber into an SX SFP - that will just require a new connector or new fiber run to the NFRA equipment racks.

CWDM has been investigated as an option to increase the bandwidth of this link to handle 1024Mb/s. But some kind of trunking would once again be needed to send traffic over both link-members. Because we control the complete path between the WSRT and us, this might be a viable option. The receiving Mark5 should either have a double 1Gb/s connection, or a 10Gb/s connection - in the first case, only one of Jive's Mark5's can be used with the WSRT, in the second case we have more flexibility but might have problems with packet reordering. A better alternative would be to upgrade the link to the WSRT itself to 10Gb/s. This would require replacing the current 1Gb/s extended range optics with their 10Gb/s counterparts, and replacing the 'convertor' switches with models which support 10Gb/s. The estimated cost of this operation at this time is approximately €20.000.

3 Requirements

These are the 'must have' requirements that were identified prior to selecting the networking equipment.

- Ports (all of these must be non-blocking)
  - 1x 10Gb/s single-mode
  - 16x 1Gb/s SX
  - 1x 1Gb/s SX or SFP (LX) for link from WSRT
  - 20x 1Gb/s 1000-baseT (RJ-45)
- Fully manageable, SNMP support for statistics
- 16Gb/s aggregate throughput (switching and routing)

Nice to have:

- Ports
  - Up to 16x 10Gb/s CX-4 or 10GbaseT (copper) to Mark5's (can be overcommitted), will need less 1Gb/s 1000-baseT.
  - 1x 10Gb/s ER for link from WSRT
  - 30x 1Gb/s 1000base-T and 3x 1Gb/s SFP (SX) additionally, to upgrade the local network in the correlator

4 Switch equipment

Several vendors and resellers of networking equipment have been contacted and invited to offer equipment for our networking upgrade. This has not been conducted as a formal RFP, but more in the way of a dialogue with their technical and marketing people. After much consideration, we have opted for an HP ProCurve 5612zl chassis, and a smaller HP ProCurve 3500yl switch for the local network. This combination fulfills 'must have' requirements at a very attractive price, and still allows us ample room to expand the network capacity.

The 5612zl is a modular switch, with 12 slots for port modules. Our initial configuration will consist of the following components:

1x	J8698A	5412zl 12-slot chassis (empty)
3x	J8712A	875W Power-supply
1x	J8437A	Port-module, 4x 10Gb/s X2 (2:1 overcommitted)
1x	J8437A	X2-SC LR 10Gb/s Optic
1x	J8702A	Port-module, 24x 10/100/1000 Mb/s RJ-45 (non-blocking)
1x	J8706A	Port-module, 24x 1Gb/s mini-GBIC (SFP) (non-blocking)
16x	J4858B	mini-GBIC (SFP) SX-lC

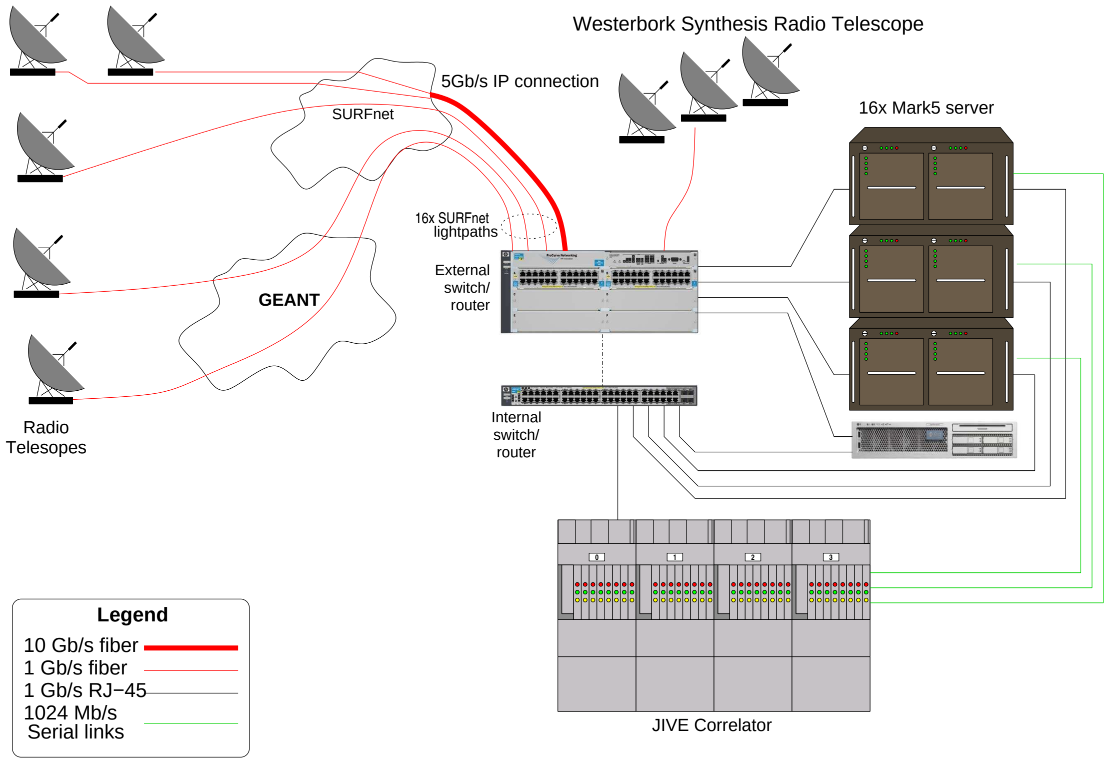

Figure 2: JIVE network after upgrade

5 Glossary

- SFP - Small Form-factor Pluggable, also known as mini-GBIC - Standard for small modular Gb/s interfaces, available with optics interfaces for multi-mode and single-mode fiber, CWDM and 1000base-T.
- XFP - 10Gb/s version of the SFP.
- CWDM - Coarse Wavelength Division Multiplexing - A relatively cheap way of allowing several wavelengths on one fiber, to enhance throughput.
- CX4 - 802.3ak standard for 10Gb/s ethernet over infiniband copper wiring
- 10Gbase-T - 802.3an-2006 standard for 10Gb/s over category 6a or 7 copper wiring, using the familiar RJ-45 connector. This standard has only very recently been ratified.