LWA Data Communications - The Sneakernet Option

Authors

Greg Taylor (UNM)

November 30, 2007
LWA Memo

Contents

1 Introduction

While the desired mode of operating the LWA is as a real-time connected element interferometer (Taylor et al. 2006), here we consider an alternative data communications plan. In the plan monitor and control is carried out over fiber or by telephone, but the data from the antennas are recorded on hard disk drives and are then carried back to a central location for correlation. This approach is referred to in general as "sneakernet". An estimate of the cost involved is presented, if only to give some comparison to the cost of bringing the data back over fiber.

The technical requirement (still to-be-confirmed) for bringing back data from each station is 1 Gbps. Here we assume that the LWA stations operate at 1 Gbps or within a factor of 2 of that on average.

Reliability of the sneakernet approach may be affected by weather and disk failures. This can be compared to a connected-element approach. Simple comparisons between the reliability of the VLA and VLBA suggest that reliability is not substantially reduced. Reliability is not addressed further in this memo.

2 Components

There are two fundamental choices to make for how the sneakernet option is implemented. These are (1) COTS (Commercial Off The Shelf); or (2) The Mark 5 recording system currently used for VLBI.

The COTS method could be less expensive but would have to be implemented by the project. Most likely this would be built on top of the modern SATA interface for disk drives. While the SATA format provides for a 3 Gbps interface, the actual sustained maximum transfer rates for most drives is about 250 Mbps, so some develepment would be needed in order to stream data out to multiple drives and achieve the required 1 Gbps recording. ETA (Ellingson, e-mail Nov 28) uses an EDT card to a SATA-based software RAID array implemented on a Linux box. This achieves 480 Mbps continuous recording. Acquisition software is written in C and runs under Linux. The estimated cost of this system (including PC) is $5k.

The Mark5A system is in use around the world for VLBI recording, and has been available since 2003. It is a commercial system developped in cooperation with MIT/Haystack observatory and sold by Conduant. This rack mountable recorder features two hot-swappable 8-drive modules, and can play or record up to 1 Gbps in its current Mark5A incarnation. The Mark5B+ system which became available in 2007 supports up to 4 Terabytes of storage in each module and delivers 2 Gbps data rates. The Mark5B system has the further advantage that no formatter is required. The Mark5C system (available in mid-2008) will support data rates up to 4 Gbps, and has the further advantage that it uses a standard 10 Gigabit Ethernet interface so that data files will appear to the user as standard Linux files that can then be read by a sofware correlator (Whitney 2007). The Mark5 modules are also readily packaged for shipment, and have understood failure modes.

Implementing the Mark5 system has the advantage that it would more readily provide compatibility with the VLBA. There is already a 74 MHz receiver system on the VLBA antenna at PT. One could imagine implementing 74 MHz receiver systems on other VLBA antennas and recording the data on Mark 5 modules for later playback and correlation with LWA stations.

In this memo we assume that the COTS recording system and the Mark 5C systems have identical functionality and cost out both.

2.1 Monitor and Control

Monitor and control (MCS) commands for the stations are naturally delivered over fiber, presumably at a negligible additional cost. Given the modest MCS requirements (48 kbps but still to-be-confirmed), a phone line should be sufficient. If fiber is not available, then installation of a phone line would be required for each station.

2.2 Time Keeping

Connected element interferometers generally maintain phase stability by distributing a common LO signal. In very long baseline interferometry (VLBI) systems where it is not practical to link the elements, data is time stamped using a Hydrogren Maser. The stability of the time standard limits the phase stability of the observations at the radio frequency. A rule-of-thumb is that one needs a time standard that is stable in frequency by a factor 20 better than the highest frequency of interest. For the LWA operating below 100 MHz, we need a time standard stable at the 2 GHz level, or $\sim$1 part in $10^{-9.3}$. GPS receivers which will be present at each LWA station for ionospheric requirements come with oven-controlled crystal oscillators which should be good to about 1 part in 10^{-11}, or about a factor 50 better than we require.

2.3 Recording

Each station would be equipped with a single Mark5C/COTS capable of recording at up to 4 Gbps. Assuming a sustained rate of 1 Gbps, the required storage is 10 TB/day. The Mark5C/COTS modules can be populated with 1 TB disks such that each module stores 8 TB, and the two modules taken together store 16 TB.

2.4 Shipping

The LWA1+ stations are located along Highway 60 and could be collected and brought to Albuquerque (or some other location between the VLA and Albuquerque) in about 1 day.

The LWIA locations are distributed roughly along a ring defined by I25, and Highways 152, 180, 12, and 60 (see Cohen & Taylor 2007). Total distance around the ring is 384 miles, about 7 hours. In practice with detours to stations, it is likely that it would take a driver 2 days to collect data from 10 outlying stations and deliver it to a central location located in Albuquerque or between the VLA and Albuquerque.

The minimum latency is likely to be 2 days to recover data from all sites. The maximum latency is set by any science requirements, and by storage limitations of the Mark5C/COTS system. Assuming that the modules are made up of 1 TB disk drives, then the storage capacity of the Mark5C/COTS is 16 TB, or about 1.6 days at 1 Gbps. If the data rate is 576 Mbps as specified in the LWA1+ system architecture document, then the storage capability of the Mark5C/COTS system will allow for 2.8 days of continuous recording. This suggests that data will need to be collected from the sites every 3 days. For full time operation each station would need 9 days worth of disk (one set being recorded, one set being delivered, one set being correlated).

2.5 Playback and Correlation

A playback unit will be required for every station. These units would feed a software correlator assumed to be capable of keeping up with 1 Gbps data in real time. Assuming a steady state model for data collection, the requirements for correlation are essentially the same for a sneakernet model as for a connected element array. A notable additional requirement is the need to keep track of disks, and for an operator to mount and unmount disks. This requires at least one full-time technician to run the correlator, whereas with a connected-element array one might consider remote operation of the array and correlator since there are no moving parts involved. We further assume that the correlator can be operated unattended for significant periods of time since the disks can play back for 3 days.

3 Cost Model

3.1 Mark5C construction costs

Mark 5C recording units: 25k/station
Mark 5C playback units: 25k/station
Disk modules needed for 9 days of operations: 48 TB, 12 k/station
assuming $0.25/GB for 1 TB disk drives in 2008 and beyond.
Installation of a phone line: 5k/station

Total cost: $67k/station

3.2 COTS construction costs

COTS recording units: 5k/station
COTS playback units: 5k/station
Disk modules needed for 9 days of operations: 48 TB, 12 k/station
assuming $0.25/GB for 1 TB disk drives in 2008 and beyond.
Installation of a phone line: 5k/station

Total cost: $32k/station

3.3 Operations costs

Below I cost out what would be needed for the LWIA. These are the same for both Mark5C and COTS. For the LWA1+ the disk replacement costs are lower, and the mileage is less than half, but these items don't dominate the cost so the result is similar.

Disk replacement cost: $1k/year/station or $16k/year.

Vehicle cost: $300/month GSA lease + 30 cents/mile. Assuming 500 miles every 3 days amounts to $1800/month including lease. This is $22k/year.

Driver cost: $40k/year including benefits
Technician/correlator operator cost: $50k/year including benefits
Annual phone bill: $4k/year for the 10 remote LWIA stations

Total cost: $132k/year for the LWIA, a bit less for the LWA1+

4 Hybrid Modes

The above analysis assumes that data from the core locations can be picked up from a single site, since these stations are all linked up by fiber. If sufficient bandwidth was available to some sites at low cost, then one could imagine shipping that data over fiber to the central facility for recording there. This could reduce the transport cost for those stations, but would still require recorders, playback units, a disk supply, truck, driver, and an operator so the cost saving for having a few stations connected by fiber is minimal. In order to realize a dramatic savings one has to have all stations available over fiber, and the fiber costs must be significantly less than the sneakernet cost presented herein.

If many sites were connected to the correlator facility by a lower bandwidth communication system (constrained by cost perhaps), then one can imagine a hybrid mode in which some high-priority projects were brought back over fiber in an "ftp" mode. That is, the data would be recorded at the sites at full bandwidth, spooled back over the network connection at a reduced bandwidth, and later correlated at full bandwidth. This capability could allow for some high priority projects to get correlated in near real-time, at the expense of the fiber use charge. Again, no savings on the cost model presented here are realized, and in fact some additional recording units might be required at the correlator.

5 Summary

A sneakernet approach to moving data from the stations to a central correlator appears to be a viable option. The cost of implementing this approach, using the soon-to-be-available Mark 5C recording system is $67k/station to implement and $132k/year to operate up to and including the LWIA. The cost of implementing this approach, using a COTS systems designed by the project is $32k/station to implement and $132k/year to operate up to and including the LWIA. Should it turn out that the cost of fiber is prohibitive for some stations, then it should be possible to implement either a COTS or a Mark5C recording system. Further studies would be necessary to choose between these two options and to come up with a design.

References

Cohen, A., & Taylor, G.B. 2007, LWA memo, in prep

Taylor, G. B., et al. 2006, LWA memo #56

Whitney, A. 2007, Mark 5 memo #57