Recording capacity research for a reliable Data Acquisition System for VLBI

Guifré Molera i Calvés
Metsähovi Radio Observatory, TKK

Abstract- This paper discusses and implementation of a high rate recording system for VLBI purpose. Using a common components from the market a low cost system has been built and tested to fulfil the EXPReS project requirements. The results demonstrate a great approach to constant recording or transfering at over 4 Gigabit/s. Furthermore, the system has still way to be upgraded and gain a margin of proper usage.

I. INTRODUCTION

In Very Long Baseline Interferometry (VLBI), radio observatories around the Earth effectively create a synthetic globe-sized giant radio telescope. The simultaneous recordings of radio signals of several telescopes located geographically apart are combined and correlated. Routine data rates in today's VLBI observations lie between 256 and 1,024 Megabits/s, with the trials already on their way at 2 and 4 Gigabits/sec (Gbps). Metsähovi Radio Observatory (MRO) focuses on developing a reliable Data Acquisition System (DAS) for recording or transferring real-time data at 4 the transfer rate of Gbps.

The objective of this investigation was to break the t data rate limit of 4 Gbps on a DAS. Even previous tests demonstrate that it is possible to achieve such rates by multi-computers or multi-RAID sets; systems have not been stable enough to use as a DAS. The new Asus L1N64-SL1 WS multitask with its 12 S-ATA integrated ports opens a new path to record data over 4 Gbps. In addition, disk could handle over 12 TeraBytes of observational data.

II. MATERIALS AND METHODS

Last december a computer, called Abidal, was built which might become a replacement for the old DAS used in the different European observatories. The Asus L1N64 was at that time the only one with an integrated 12 S-ATA ports, and it also supported two dual-core processors. The benchmark of the computer was not a disappointment. In addition the new Samsung 750 F1 SpinPoint HD753LJ [2] were also bought earlier 2008. The newest Samsung hard drives are far ahead from the rest of competitors in write and read capabilities. Regarding the disk benchmarking, it is commonly extended the use of XFS as a File System format in our tests due to faster write & read operability rather than other common linux FS. [3]

The different set of tools used to benchmark the system and the performance of the disks are both developed at the laboratory: Wr and Tsunami-UDP. Firstly, Wr is a software tool developed at MRO which reads data from a VSIB board[1] used in the VLBI observations to record data on Commercial Off The Shelf (COTS) computers and write it to disk.

Secondly, Tsunami is a fast aggressive FTP, which can also be used as a data transmission protocol. Tsunami was developed at Pervasive Technology Labs research center at the University of Indiana and currently is developed completely at MRO. Transmission data is send as UDP/IP packets whose transfer rate is significantly higher than that of TCP, especially with long distances. The basic idea behind Tsunami is that the transmission data is chopped to large packets of equal size. The goal transfer rate is 650 Mbps.

Finally, a second test computer was configured to send dummy data by using Tsunami-UDP server to our recording system through a direct fiber connection. Secondly, Abidal was set to handle each tsunami transfer in a different processor, to optimize both cores to gather data packets and write them to disk.

A summary of the tests performed with the new system is as follows:

- ✔ Multi-thread recording to single disk: Using Wr software, data can be recorded to different directories, hard disks or RAID systems simultaneously. Each new allocation is handled by a different thread.
- ✔ Single-thread recording to RAID system: Using Wr, data is recorded to a single RAID disk in

order to fill all its capacity depending the amount of disks used.

- ✔ Multi-thread recording to RAID system: Using Wr, data is recorded in different RAID locations handled each one by own thread.
- ✔ Tsunami-UDP client. Using Tsunami-UDP transfers data sended by other pc can be received and recorded into the disks. Simultaneous calls to the program were used in order to run two or three Tsunami-UDP streams in parallel and to achieve higher rates.

III. RESULTS

A- Multi-thread recording to single disk

By several processes simultaneously writing rates over 5.5 Gbps to single disk is achievable when the CPU is not overloaded. While a maximum rate for a single S-ATA disk is limited to 711 Mbps, the coefficient between total throughput and number of disk keeps constantly at 700 Mbps up to 8 disks. Unfortunately, the usage of higher amount of disks just reports a slight improvement in the total rate. Even when the threads share the process within 4 cores, those get totally overloaded when the twelve disks are used. Consequently, the rate by using 8 disks does not differ barely from using all 12. Figure 1 illustrates the evolution of recording rates on simultaneous SATA disks.

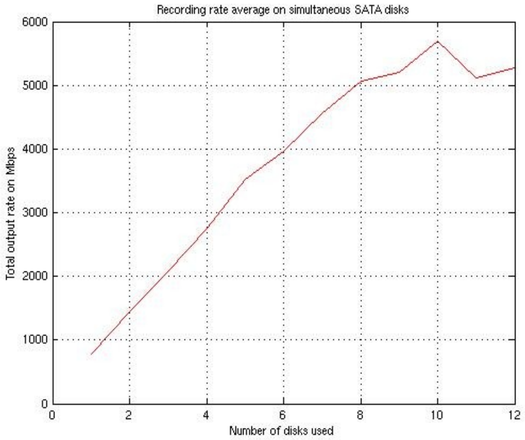

Figure 1: Total output rate recording while several single SATA disks and multi-threads are used. The writing average keeps constant until eight disks.

B- Single-thread recording to RAID system

While the CPU load limitation is less noticeable when several disks are handled as a big single unit, the recording rate does have lower values than in previous experiment. As it is well known, the performance of a RAID disk is always a bit worst than the same amount of separate disks. On the other hand, when a set of all 12 disks is in use the rate achieves its top (5.5 Gbps) mainly due to the fact that the CPU load is not mainly conditioned by the size of the RAID system. Figure 2 shows a comparison between writing rates and disk space usage while using 8,10 and 12 disks on a RAID system.

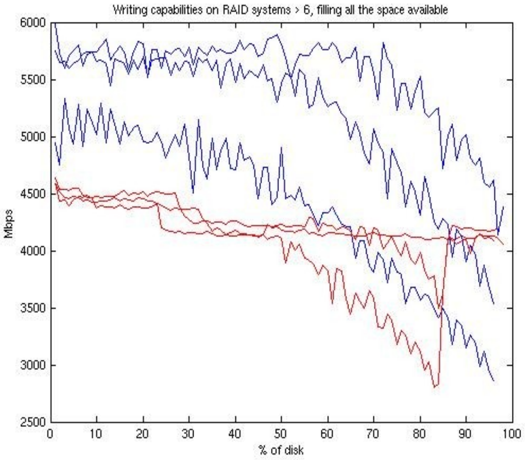

Figure 2: Writing capabilities on RAID systems depending on the disk space usage. The RAID's are mounted with 12,10 and 8. Also in the plot the lines are distributed in same order(higher/lower). Blue lines are for raw disks and red ones used XFS format.

On the other hand, the CPU constraints appear when the raw RAID disks is substitued by a formatted File System. As told previously, XFS is used as FS disk format. The limitation become clear when a high amount of disks, over 6, is used to create the RAID. Figure 3 is a comparison of using raw (blue) and XFS (red) disk format. Even at a first glance we believed that File System does not have a strong influence in the behaviour of the disks we realized that the usage of any kind of File System has an effect of the amount of resources used by the processor. A reduction of 25 % of the best performance occured while while using all 12 disks. In any case, by using a File System it seems still possible to record data over 4 Gbps, but the margin we are working with it is quite tight to ensure the transmissions will work stable and without trouble. Note that in Figure 2, a constant recording rate of over 4200 Mbps, which lasted throughput the whole test (5 hours), was achieved.

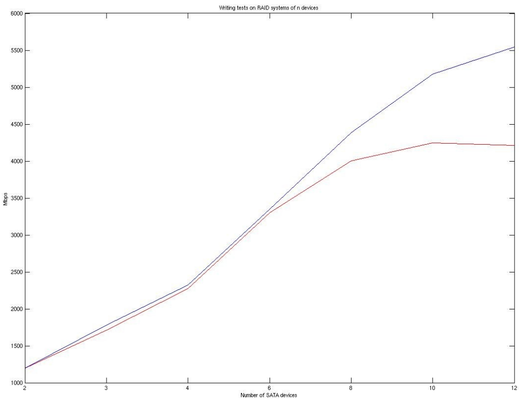

Figure 3: Comparison of writing rate by using a RAID system disk depending on the number of the disks. The RAID format was raw (in blue) and XFS (in red).

C- Multi-thread recording to RAID systems

Undoubtedly, combining parallel thread and RAID disks to optimize both, total CPU consumption and single core usage of each processor, showed the best results. The basis is to optimize each core of the dual processor to handle the writing process into the RAID disks itself. Obviously, the maximum number of RAID's is limited to the amount of available cores, 4. Three different tests were done:

- 4 RAID's 3 disks in each. All 4 cores in full use.
- 3 RAID's 4 disks in each. 3 main cores in use.
- 2 RAID's 6 disks in each. 1 processor for each RAID.

	raw format		XFS format	
	indiv. G	total M	indiv. G	total M
2 RAID	3.9/4.1	8023	x	x
3 RAID	2.5/2.5/2.4	7559	x	x
4 RAID	2/1.9/2/2.1	8150	x	x

Using parallel tasks in several several RAID systems is a great improvement which optimizes the processes on the process side. That is really important when the data will be supplied to the system through the 10 Gigabit ethernet link and will require special criteria for handling the incoming packets and so an extra use of CPU.

D- Tsunami-UDP client tests

Sustained transfer/recording rates over 4.7 Gbps were easily achieved by creating different Tsunami client instances on the pc. Logically, by using 2 RAID and 2 Tsunami instances the best output was achieved. Each processor can manage each transfer correctly and the tasks of obtaining packets and writing them are shared between both cores. Even though, when a more complex configuration is used, no higher than 3 Tsunami parallel streams, the system still improves the performance of a single thread. The following Table 2 shows a comparison of the results.

	XFS format	
	individual (Mbps)	total (Mbps)
1 RAID x 12	4350	4350
2 RAID x 6	2609 2625	5234
3 RAID x 4	2025 1409 1321	4755
4 RAID x 3	xx	xx

When 3 Tsunami threads are used in the tests the result demonstrate a kind of master/slave behaviour between the transfers. One of them is the master with a rate over 2 Gbps and the other two are slaves with values around 1400 Mbps. A couple of reasons could be the answer to the problem: one of processor arranges one stream, and the other two are managed in the second CPU. On the other hand, the second computer used is less powerful and all the limits seen on the tests could be caused from it.

E- Tunning few disk and network paramters

Linux offers an infinite amount of small tricks to improve the perfor-mance of the computer, unfortunately, the gain achieved is really low at those high rates. A recommended modification is listed:

- Usage of MTU blocksize set always to 8192
- Usage of large UDP packets, commonly 32 kbytes.
- Optimize TCP and UDP windows size parameters on both sender and receiver computers.
- High priority enabled do not maximze the use of Tsunami.

IV. DISCUSSION

Even using single or multiple parallel tasks or using different disk distribution the CPU consumption limits the capability to record data into the disks. Hence, the AMD quad-core processors might give a helpful hand to solve processor problems. The price of them are totally out of range for our budget and specially if the main purpose is just testing. Even though, the components prices are constantly dropping and it is not a crazy idea to reckon that at the end of year could be reasonable to purchase them.

On the other hand, the specifications fo the disks used during the tests, Samsung 750 F1, promised to have a performance almost twice better than the old ones. Tests demonstated that even the the improvement was over 10 % was far from the hoped one.

Regarding the tests, combining multi parallel tasks and recording to various RAID's over XFS format has not been concluded due to lack of time. Our best guess will be a decrease on the performance around 20 % as seen previously. In fact, those 6000 Mbps might still become a very optimistic result. Nevertheless, currently the effort is focused on optimizing the usage of the tasks on the processor side and that points to test's section D. The software department is working to improve the current source code to automate multi-thread for itself.

Finally, the second test machine is a dual-core processor and is already a year an half old. The CPU was loaded in all the tests up to 99% of its possibilites. In a near future, we hope to repeat the tests by using a better computer to ensure that is not the limiting factor.

V. CONCLUSIONS

While the project in which we are involved, EXPReS is still in development at least until middle of 2009. Our goal to achieve a reliable Data Acquisition System at 4 Gbps is almost done. Thanks to the constant upgrade of the new COTS components turns us really optimistic to have a ready and stable system at the ends of years, at a relative low-cost price.

VI. BIBLIOGRAPHY

[0] EXPReS project – http://www.expres-eu.org
[1] VLBI standards and specifications – http://www.metsahovi.fi/en/vlbi/
[2] Samsung 750 GB disk benchmarking - http://tweakers.net/benchdb/testcombo/1517
[3] Disk and File System benchmarking at Metsähovi - http://www.metsahovi.fi/en/vlbi/10gbps/10gbNetworkTests.html