The ANTARES Offshore Data Acquisition: A Highly Distributed, Embedded and COTS-based System

Shebli Anvar[*], Hervé Le Provost[†], Frédéric Louis[‡]
DAPNIA/SEI – CEA Saclay – F-91191 Gif, France
[*]S.Anvar@cea.fr, [†]H.LeProvost@cea.fr, [‡]F.Louis@cea.fr

Abstract

ANTARES is a neutrino telescope project based on strings of Cerenkov detectors in deep sea. Some 400 detectors are spread over a volume of 3 000 000 m^3 at 2000 m underwater, which calls for a highly distributed, embedded data acquisition system. Based on Comodity Off-The-Shelf (COTS) low-power integrated processors and industry standards such as Ethernet and CORBA, the foreseen system is expected to run for 3 years without any direct access to the hardware. We present the system's performance needs and constraints, and discuss the solutions that were imagined. We also make a case on the advantages and drawbacks of COTS electronics and software, and present recommendations on how to use COTS components efficiently in similar projects and HEP projects in general.

I. INTRODUCTION

The ANTARES telescope is designed to serve as a cosmic neutrino detector for astroparticle research [1]. The idea is to detect Cerenkov light emitted by the superluminic muon that is created through the interaction of any high-energy neutrino with matter and reconstruct the energy and the trajectory of the neutrino. The system must be wide enough to be able to reconstruct trajectories with sub-degree accuracy. As other neutrino experiments [7] [8] [9], the ANTARES detector must reside in a transparent material medium whose refraction index allows for the emission of Cerenkov light, namely solid or liquid water. At he same time, in order for the medium to shield the system from atmospheric muons, the ANTARES detector will reside at 2000 meters under the Mediterranean sea. The detection of the Cerenkov light is achieved through photomultipliers sampled by arrays of a specific ASIC called the ARS1 [2]. The so-called "Offshore DAQ" is the acquisition system in charge of reading out the digitized data produced by these ASICs throughout the whole detector and transmitting them to a computing farm on the shore through fibers inside a submarine cable. The offshore system presents performance challenges together with specific constraints due to its situation at 2000 m under the sea level.

Section II is devoted to the description of the main requirements and constraints of the offshore DAQ and the type of solution this entails. In section III, the DAQ boards and the functions they achieve are described, together with the DAQ network topology. Section IV is more specifically devoted to the offshore software and its design principles. In section V, as a conclusion, we propose to extend the principles we have used in the design of the ANTARES DAQ system to HEP and "big science" instrumentation systems in general.

II. CONSTRAINTS AND REQUIREMENTS

A. The Physical Distribution of Detectors

The detector is made of 13 strings of 30 detecting nodes each. The strings are spread over an area of 0.1 km². A detecting node consists in 3 photomultipliers each sampled by an array of 2 ARS1 ASICS. On each string, there is a detecting node every 10 m. Each detecting node is equipped with an electronics crate called the Local Control Module (LCM), in which the DAQ boards (amongst others) are situated. The main function of DAQ boards is to readout the 3 pairs of ARS1 associated with the detecting node and transfer the data to the shore workstations. It may process the data to some extent. It also runs the offshore slow control software through which all the electronics of the LCM crate is controlled.

B. Constraints And Requirements

1) Cable Flexibility Vs Reliability

The topology of the offshore part of the network is a compromise resulting from constraints on reliability, robustness and mechanical space: a daisy chain minimizes space occupation (fewer cables) but knocks down reliability as the failure of one node in the chain would result in the loss of at least all higher nodes. Consequently, a full star topology would be preferable in terms of robustness and reliability, but such a scheme would entail too thick and rigid an electro-mechanical cable (EMC): the EMC must be flexible enough for the string to be stored on boat for sea operations. The right compromise is therefore a tree topology: each string is considered as a tree of 6 substrings of 5 nodes each (Fig. 1). A substring is called a "sector".

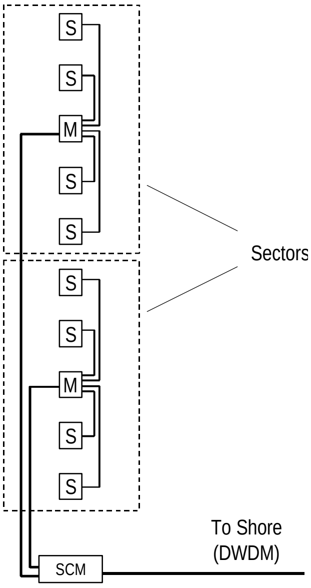

Figure 1: String Sectors

The data-flow coming from the nodes of one sector is concentrated in one of the sector nodes called the “Master LCM”. From that point on, the data is transmitted through one fiber to the bottom of the string and then to the shore stations through “Dense Wavelength Division Multiplexing” (DWDM) as if on a dedicated single fiber.

2) Bandwidth Constraints

According to simulations and measurements, total average data flow through one LCM (i.e. 3 optical modules sampled by 6 ARS1 chips) is expected to be as much as ~25 Mb/s. Moreover, due to many factors such as bioluminescence, this data rate is expected to undergo large fluctuations leading to peak rates of up to ~120 Mb/s. The system must therefore be able to absorb such fluctuations while sustaining the average throughput. If we consider today’s COTS network components, we are naturally led to use 100 Mb/s Ethernet on each LCM for data output.

3) Space And Power Consumption Limitations

The LCM cylindrical container includes a custom electronic crate featuring circular board of ~12 cm in diameter. Interconnection between boards is made through a custom backplane. The crate will contain about 15 boards, including (e.g. ARS1 boards, acoustic positioning, power, etc.) The overall available power is ~30 W. Moreover, since there cannot be any ventilation system, heat dissipation takes place through conduction only. The LCM container is made of titanium, which is a poor thermal conductor. Since the MTBF of electronic components is an decreasing function of temperature, it is therefore imperative that heat dissipation is kept at a minimum. Space limitation calls for as much integration of components as possible, which is also favorable for dissipation minimization. This consideration has led us to the use of embedded processors that integrate many I/O capabilities. These capabilities make it possible for the processor to also carry out slow control activities. As a consequence, the main DAQ board will also support the LCM slow control.

III. DAQ Boards And Network Topology

A. The Main Board

As stated in the previous section, the main DAQ board must carry out the tasks of both DAQ and slow control. The design uses an embedded, low-consumption processor of the PowerPC MPC860 family [4] for software and communication tasks, leaving to a firmware component of the ALTERA Apex family the readout of the ARS1 digitizer chips. The MPC860P runs at a maximum clock of 80 MHz and dissipates about 700 mW. It integrates one 100 Mb/s Ethernet port, four other serial I/O ports and a bus arbiter. The slow control task uses the serial ports to carry out all the management of the LCM boards through the specific backplane. The whole detector features 13 lines of 30 LCMs, which means that the ANTARES 0.1 km^2 project will include ~400 underwater embedded processing nodes.

The data coming from the six ARS1 chips of the LCM is readout, partially formatted and stored in the board’s SDRAM by the FPGA. A small amount (a few MB) of the board’s 64 MB of SDRAM is used to store the real-time operating system (RTOS) and the DAQ and slow control applications; the remaining memory space is used to absorb the data flow. The memory bandwidth is over 100 MB/s Together, the array of ARS1 chips can produce data at a maximum rate of 120 Mb/s; the SDRAM can therefore store at least 4 seconds of data coming from a high intensity burst (due, for example, to bioluminescence). The whole main board is expected to feature less than 5 W power consumption and 300 000 hours of MTBF.

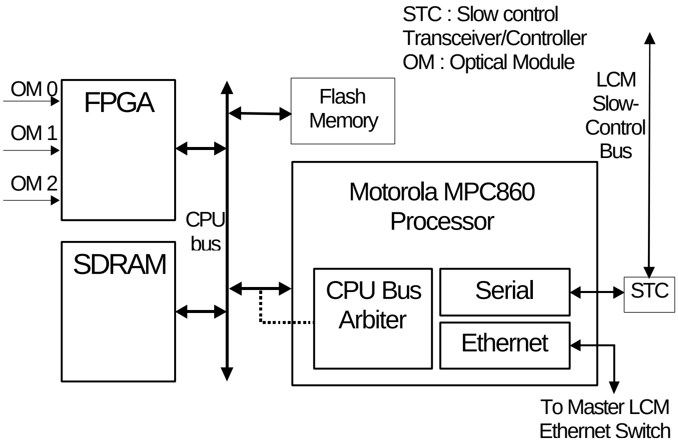

Figure 2: Main Board Block Diagram

B. The Switch Board

As described in section I.B.1, the data of each sector is concentrated in the so-called “master LCM” before being set to shore through DWDM. The data concentrator is based on two COTS Ethernet switch chips: the Allayer AL121 and AL1022 [10], [11]. The switch board will thus feature 5 Fast Ethernet 100 Mb/s ports and one 1000 Mb/s ports,[1] allowing for the transparent routing of the data coming from the five 100 Mb/s ports of a sector. The typical dissipation of this board is expected to be less than 8 W.

C. Network Topology

The ANTARES detector is made of a great number of detecting nodes spread over a large geographical area (in this case, a volume of ~30000000 m^3). A physical event is defined as light detected in synchronized time windows, which means that logically, the data corresponding to one event is originally spread out over the detector and must be grouped together to form one complete event packet. This regrouping is called the “event building,” and just like most HEP

[1] The chips allow a maximum of eight 100 Mb/s and two 1000 Mb/s ports.

experiments, the ANTARES DAQ system must achieve such an event building. The rebuilt event must then undergo a “trigger,” i.e. a real-time analysis that determines whether the corresponding data is of physical interest (and should therefore be stored on tape) or not (and should be dumped).

1) The Onshore Processing Farm and the Network Structure

The expected total data flow (~7 Gb/s) cannot be coped with using a single computer. Consequently, a processor farm of about 100 Linux PCs with Gigabit Ethernet connection will handle the onshore processing. Therefore, as shown in Fig.3, a full 78x100 switch must route the data coming from the 78 sectors’ Gigabit Ethernet connections to the farm workstations.

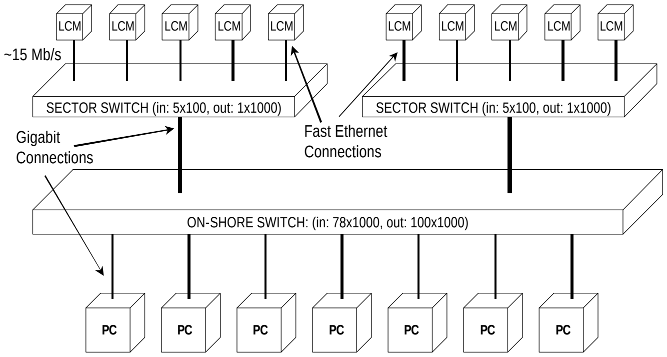
LCM
~15 Mb/s
SECTOR SWITCH (in: 5x100, out: 1x1000)
SECTOR SWITCH (in: 5x100, out: 1x1000)
Gigabit Connections
Fast Ethernet Connections
ON-SHORE SWITCH: (in: 78x1000, out: 100x1000)
PC

Figure 3: Network Architecture

2) Intelligent DAQ Processing to Avoid Congestion

The main weakness of such network architecture resides at the output ports of the sector switches. Indeed, since the sector switches are used as data concentrators (to minimize the number of fibers) an overflow of a Gigabit output port due to the communication between *one* LCM-PC pair is liable to block all communications between the LCMs of the same sector and the PC farm. This type of congestion is commonly known as “head-of-line blocking.” We must therefore guarantee that no backpressure due to an LCM-PC communication will ever propagate through the switches. This entails that the data flow at the output of each LCM must be controlled and monitored, which is another reason for having some processing intelligence inside each LCM node.

IV. The Off-Shore DAQ Software

With ~400 offshore processing nodes and ~100 onshore workstations, the ANTARES Trigger/DAQ application is clearly a massively distributed system. To enforce a coherent development and produce intelligible and maintainable software, we must adopt a methodology that encourages modularity and separation of concerns, hence our decision to design the system using the object paradigm and base as much as possible the specification of interfaces and architecture on industry standards. The programming languages used are C++ for the real-time software and Java for less time-critical code. As a matter of fact, we contemplated the use of Java even for the real-time code but we eventually considered that it has not yet achieved enough maturity in the real-time market.

To tackle the distribution problem and remain as compliant as possible with industry standards, we intend to use a CORBA middleware, both for slow-control and data acquisition. However, a data acquisition system being a real-time system, the middleware must have a dynamic behavior that is compatible with a real-time system. Such CORBA middleware exists, an example being TAO, an open-source CORBA platform featuring real-time capabilities (including the CORBA 3.0 real-time additions) [3], [4].

Separating concerns also allows for specific optimization of critical modules. Contrary to the widespread belief that modularity decreases performance, studies have shown [3] that complex systems that have a modular design are also those that display the best performance, because modularity allows for specific, localized optimization.

The robustness and overall quality of the software system will depend on the modularity of the architecture. As the system will be remote, late debugging will be at least unpractical, and the important size of the system will not allow for frequent resets without entailing non-negligible time costs. The offshore software will therefore have to be very reliable and almost never necessitate hardware resets. Apart from intensive testing before deployment, such reliability standards can be met through a careful design of the software architecture.

To support a global understanding of the software and the development of simple and modular patterns without unnecessary code details getting in the way of intelligibility, we propone the use of the Unified Modeling Language (UML), which is today’s industry standard notation used for architecture design, design pattern description and system specification. Moreover, we have begun a more general reflection (materialized as a design framework) on design and distribution patterns in HEP Trigger/DAQ system design [12]. We have already fleshed out generic concepts and defined some domain-specific UML extensions concerning separation of concerns, modularity and software reuse in distribution and hardware/software frontier specification. The design of the ANTARES offshore DAQ will benefit from the use of these concepts. To enforce as rigorously as possible the conformity of the code to the UML specification, a UML code generation tool will be used [12].

V. Conclusion: System Design for HEP

The ANTARES offshore DAQ is an illustration of the difficulties encountered in HEP Trigger/DAQ design. It cumulates massive distribution, embedded constraints, large data flow and extensive real-time processing. It is an occasion to put into practice a number of design principles that we believe as being good practice. The core of our methodology is 1) the design of modular architectures through separation of concerns, and 2) the maximization of maintainability through the use of industry standards and non-proprietary COTS.

A. Separation of Concerns

We believe that a quality design practice should explicitly distinguish functional system specifications from implementation decisions. For instance, how the system is deployed, what is implemented in hardware and how objects communicate within the system must be separate questions from what functions the system must ultimately achieve [12]. To actually enforce such a design practice without imposing awkward procedures to designers and developers, we recommend 1) the use of industry standard notations and languages such as the UML and 2) the use of automatic model transformation and code generation tools.

B. Non-proprietary COTS

It is well know that cost-effective design and maintenance avoids re-inventing the wheel through the use of COTS software and hardware components. However, an indiscriminate use of COTS might lead to impasses and become an obstacle to maintainability especially concerning software. It is particularly the case when COTS means dependency towards a specific vendor. HEP experiments are long-term projects that are bound to make use of the maintenance and evolution efforts of the electronics and software industry, they cannot afford to see their evolution disrupted because a vendor loses interest or ceases to exist. Source code availability, for instance, can prevent COTS software from becoming an obstacle to evolution and optimization, and the present trend towards open source software is, we believe, a positive direction.

VI. References

[1] The ANTARES Collaboration, "A Deep Sea Telescope for High Energy Neutrinos - Proposal for a 0.1 km^2 detector," May 1999, obtained at website http://antares.in2p3.fr

[2] D. Lachartre, F. Feinstein, "ASICs for ANTARES Offshore Electronics," to be published by *NIM*, presently at http://antares2.in2p3.fr/Publications/conferences/2000/unreg_Lachartre_1999.pdf

[3] D. Schmidt, D. Levine, C. Cleeland, "Architectures and Patterns for Developing High-Performance, Real-time ORB Endsystems," *Advances in Computers*, Academic Press, 1999.

[4] D. Schmidt, D. Levine and S. Mungee, "The Design of TAO Real-Time Object Request Broker," *Computer Communications*, vol. 21, n°4, Elsevier Science, 1998.

[5] Motorola, "MPC860 PowerQUICC User's Manual 07/98 REV.1," obtained at website http://www.mot.com/SPS/RISC/netcomm/docs/pubs

[6] Altera, "APEX20K Programmable Logic Device Family Data Sheet March 2000 ver 2.06," obtained at website http://www.altera.com/html/literature/la20k.html

[7] A. Karle et al., "Observation of high energy atmospheric neutrinos with AMANDA," obtained at website http://area51.berkeley.edu/manuscripts

[8] I.A. Belolaptikov et al., "The Lake Baikal Deep Underwater Detector," *Nuclear Physics B* (Proc. Suppl.), vol. 19B, p.388, 1991.

[9] I.A. Belolaptikov et al., "The Lake Baikal Neutrino Project ," *Proceedings of the Third International Workshop on Neutrino Telescopes*, p.365, Venice Italy, 1991.

[10] Allayer, "AL121 8-port RoX-II Switch IC Product Brief," obtained at website http://www.allayer.com/products.html#Datasheets

[11] Allayer, "AL1022 Dual-port Gigabit Ethernet Switch IC Product Brief," obtained at website http://www.allayer.com/products.html#Datasheets

[12] S. Anvar, F. Terrier, "A Design Framework for Distributed Data Acquisition and Triggering Systems in High Energy Physics Experiments," accepted for publication in the *Proceeding of the 2000 IEEE Nuclear Science Symposium and Medical Imaging Conference*," October 2000.