A Storage Network for Inexpensive RISC Workstations

Bill Joy

August 20, 1985

ABSTRACT

A rather intimidating change in technology for workstation networks is about to take place, changing all the relevant constants by an order of magnitude, while the cost stays the same:

* CPU's will increase in power to 10 times the Sun-2 class machines with the RISC by mid-1986.
* 100 Megabit local area networks, which are 10 times faster than Ethernet, should appear.
* Very inexpensive Megabit DRAM chips should be available. By mid 1987 these chips should sell for on the order of $5, making a Megabyte of memory list for $200.
* Finally, dense and inexpensive optical disk storage should appear. Estimates are that a 500 megabyte optical disk with current 5 1/4" disk performance should cost $500 in 1987.

We propose a set of workstations and storage servers using the new component technologies.

Description of the proposed network

We consider shared information to be the key to a workstation environment, and describe how a network of nodes looks, with a centralized file server.

The workstations in the network are 10 MIP Sunrise architecture with 16 Megabytes of memory. Each workstation is a single board machine with the Ferrari-style architecture, supporting a 1600*1280 black and while or 1152*900 color display. There is local support for P2IO devices.

Workstations are connected on 100Mbit fiber optic networks (ala AMD). Incoming packets are received into a 1 Megabyte RAM buffer which supports static RAM access speeds, and can DMA into main memory at very high speed. They are DMA'd to the main memory under host control after appropriate low-level processing.

At the other end of the 100Mbit fiber optic network is a RISC-based file server processor. It receives packets into a circular buffer, like the workstations have. It can process read-requests directly, by accessing a DRAM large cache memory. Reads over the network are thus normally memory to memory transfers. It processes *write* requests by placing the request in a CMOS static memory protected with ECC (after making appropriate checks), and responds when this memory is committed before doing the actual disk input/output. This allows burst data transfers to begin before the disks are on-cylinder.

The server uses multiple read-write optical disks to increase its bandwidth and to decrease latency. It can also be slaved to a second processor to make it redundant. This second processor can be used to process half of the memory cache misses which the first processor sees, so that the effective cache memory is doubled when both the main processor and the hot spare are in use.

Note: the VAXcluster hardware and software for VAX/VMS is similar to this in spirit, just fantastically more expensive.

First product

The first product uses the Ferrari board and Ethernet. We will implement a P2IO device with CMOS memory and a battery to protect the log. This can be 1 Megabyte built of 64K CMOS SRAMs and a battery and put on the "paddle" of a P2IO expansion board. Mass storage for this system will be high-density 5 1/4" disk drives, accessed over SCSI.

The main memory of Ferrari (upto 16 Megabytes) should make an adequate cache.

A special P2IO modules connects two of these servers to make a ultra-reliable server.

Cost target: $7,900 for 4 Megabyte b/w diskless Ferrari, $9,900 for 8 Megabyte cache server with everything but the disks.

Performance target: An aggressive goal for this server would be to serve clients hitting the Ethernet with about 750 kilobytes per second. This assumes an extremely large number of read requests satisfied in the memory of the server.

Second product

The second product uses the SCRAM cache version of Sunrise as the server board. A second generation "paddle" should support 2 or 4 Megabytes of log built of 256K CMOS SRAMs and a battery. If the base SCRAM Sunrise board does not support the 100Mbit local network, then it can be interfaced with a second paddle.

For this product we will also build a P2IO module to control 5 1/4" optical disk drives; a large server would use several of these to get very high data rates.

Cost target: $9,900 for 16 Megabyte b/w diskless SCRAM machine $9,900 for 16 Megabyte cache server with everything but the disks. Disks should be $2,000 or less per 500 megabytes. A performance target would be 3 megabytes per second to clients with a 4 gigabyte server (8 disk arms), in a server costing $20,000. This server costs more than $200,000 today.