An Introduction to Network-Attached Storage Devices: CMU's Perspective

Garth Gibson, David Nagle
Computer Science and Computer Engineering, CMU

Cost-effective scalable bandwidth

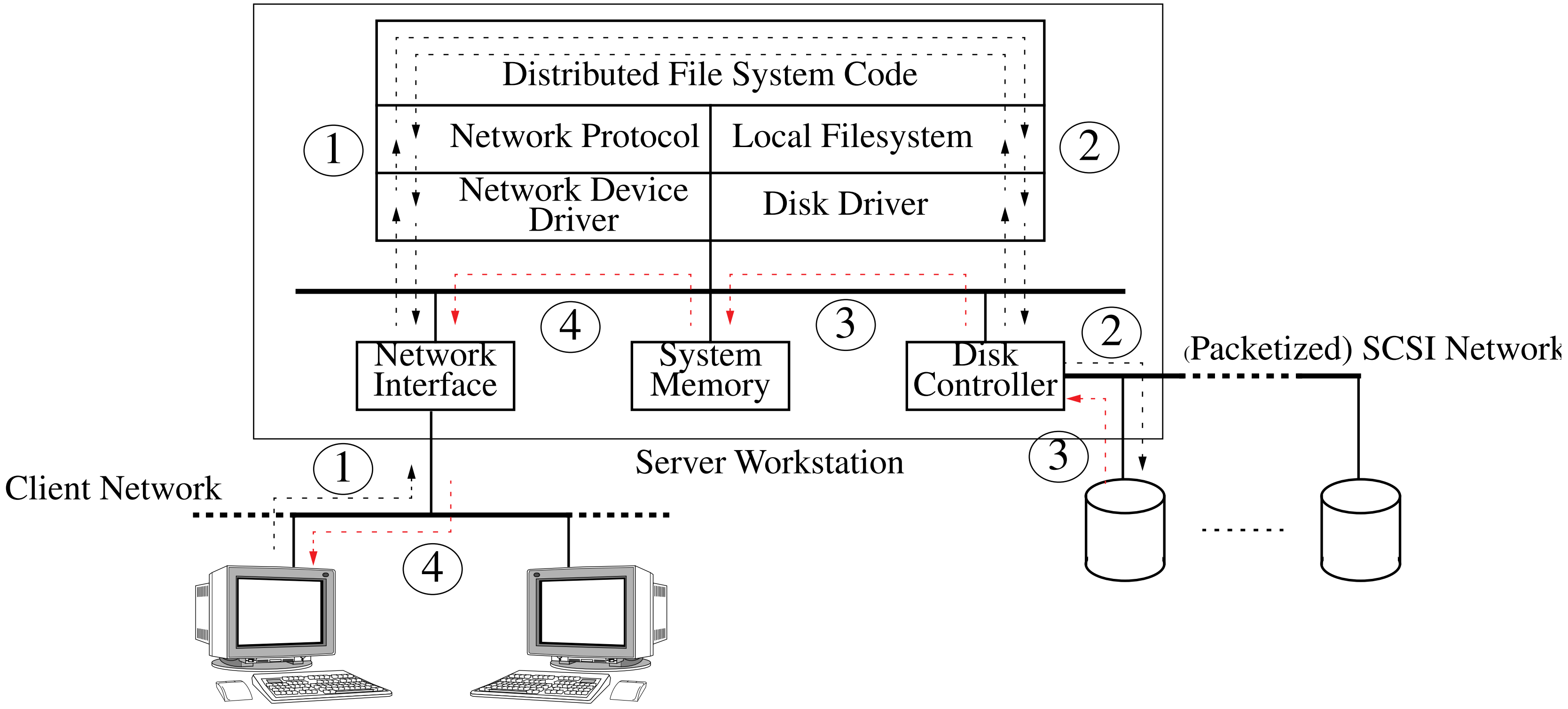
Wire-once infrastructure for storage, cluster & LAN

Sponsored by DARPA/ITO Quorum/Scalable Systems
and HP, Quantum, Seagate, STK, Symbios, Clariion, Compaq, Wind River, Intel, 3Com

Consider our current Server-Attached Disk

Store-and-forward data copying thru server machine

- **translate and forward request, store and forward data**



Carnegie Mellon

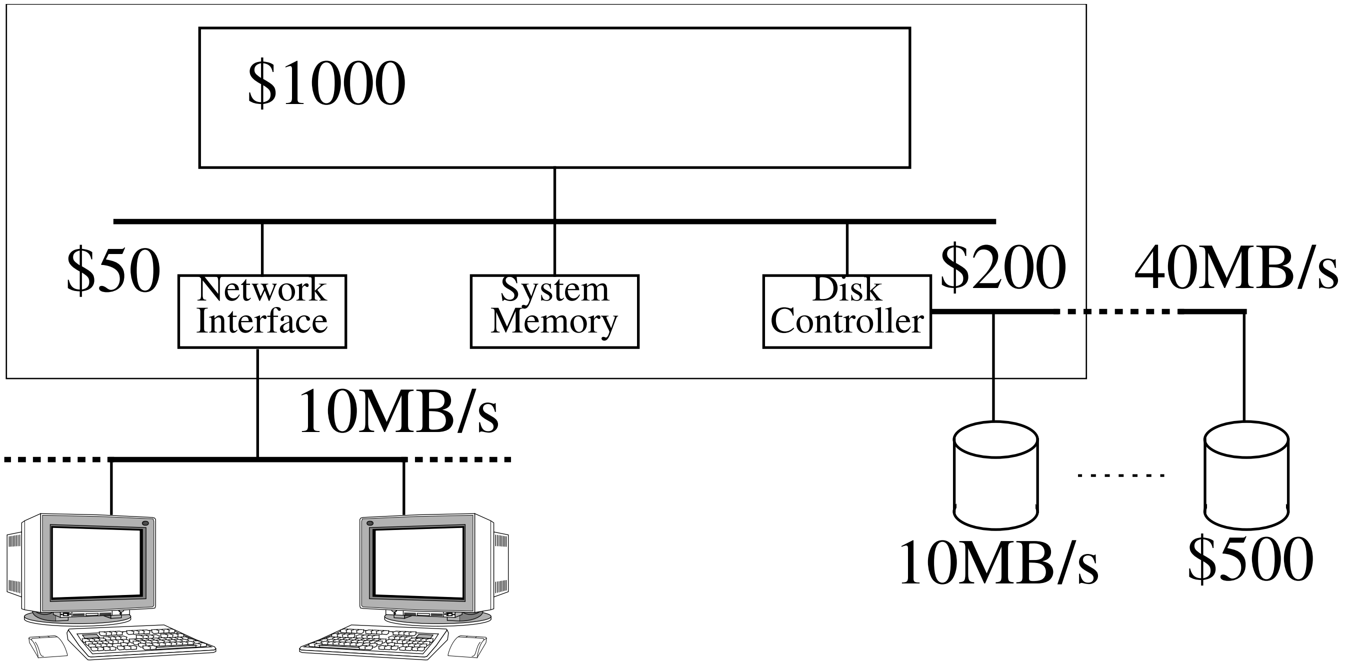
Parallel Data Laboratory

Lets put some numbers on it

Cheap server workstation, 100Mb ether, UltraSCSI

- don’t ask if server has cycles, PCI bandwidth or PCI slots
- one net, one drive with server overhead cost of > 200%
- AMORTIZE:

6 nets/drives = 50% overhead;
12 nets/drives = 35% overhead;
min overhead is 20%



The Fix: Partition traditional distributed file server

Enable direct transfer between client & storage device

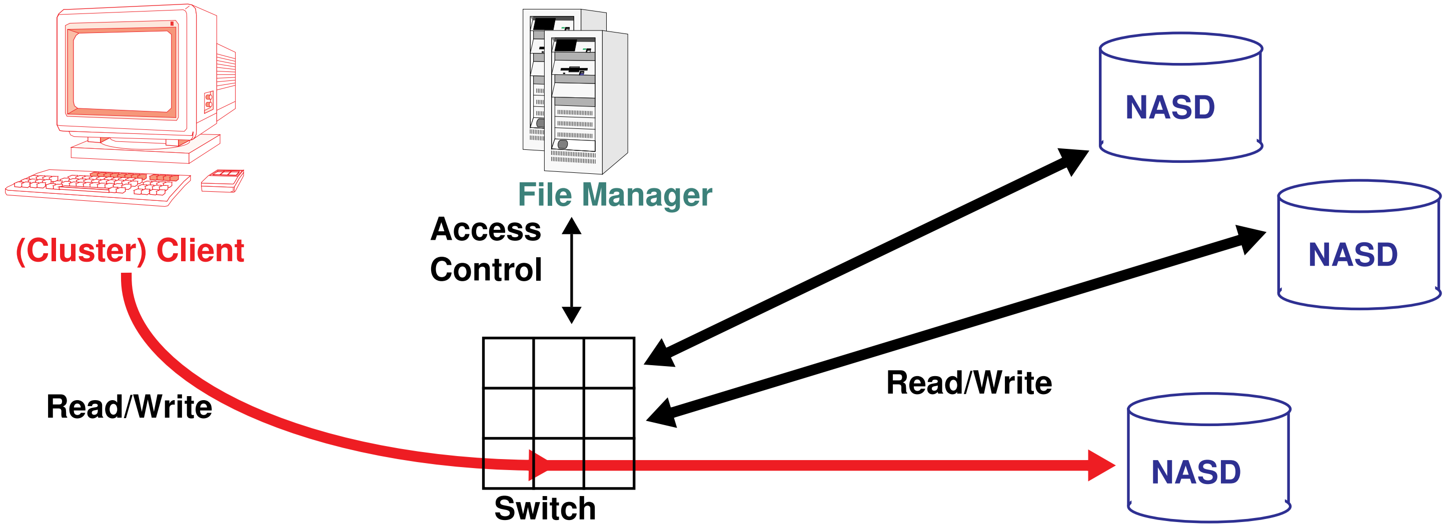
Low-level networked storage device

- direct read/write, high bandwidth transfer

Policy moved to file manager

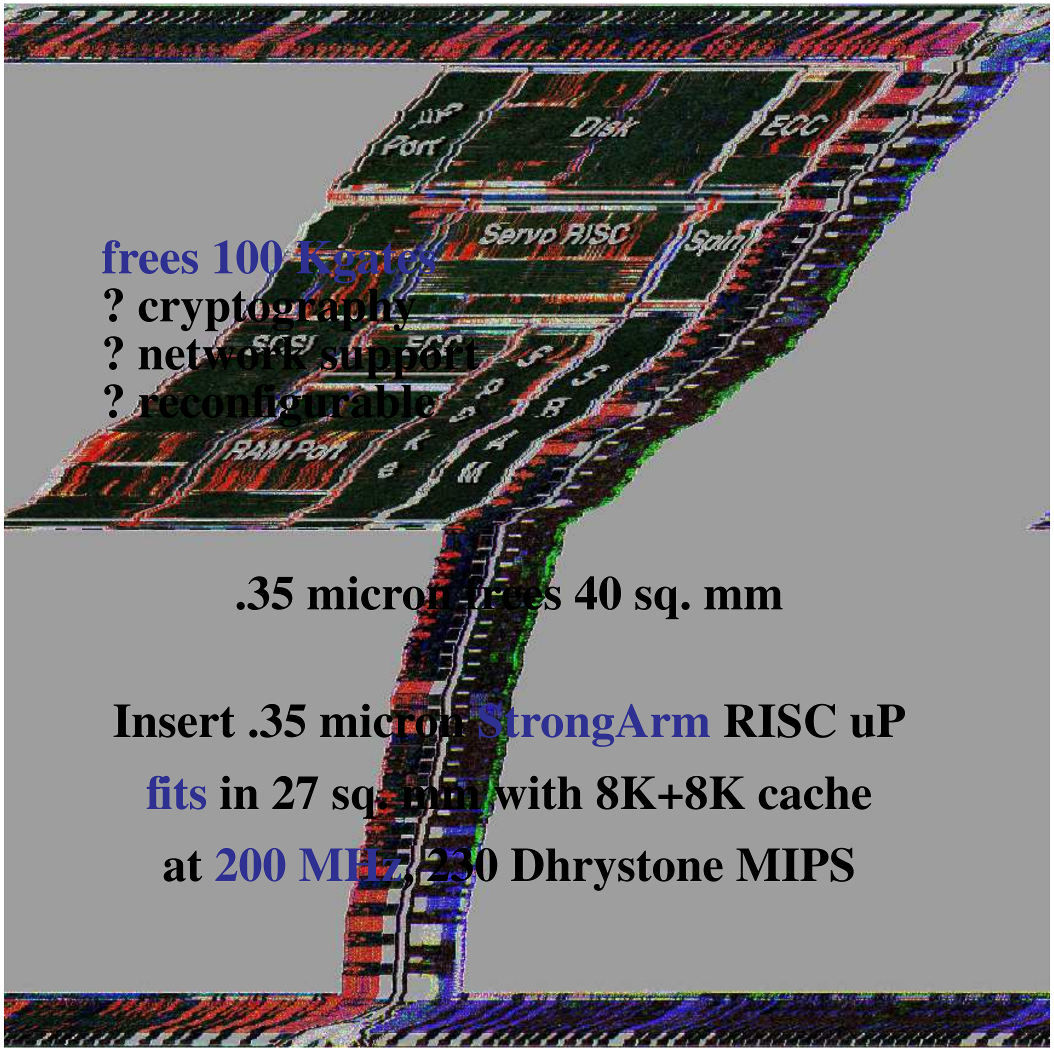
- naming, access control, consistency, atomicity

One part of NASD project – develop “right” interface



Are Device Cycles Really Available?

Current .68 micron chip is 74 sq. mm



frees 100 Kgates
? cryptography
? network support
? reconfigurable

.35 micron frees 40 sq. mm

Insert .35 micron StrongArm RISC uP

fits in 27 sq. mm with 8K+8K cache

at 200 MHz, 230 Dhrystone MIPS

SCSI processor
disk R/W channel
uP control port
DRAM port

- **~ 110 Kgates + 22Kb**
- **.35 micron next gen. enables integration of control uP onto ASIC**

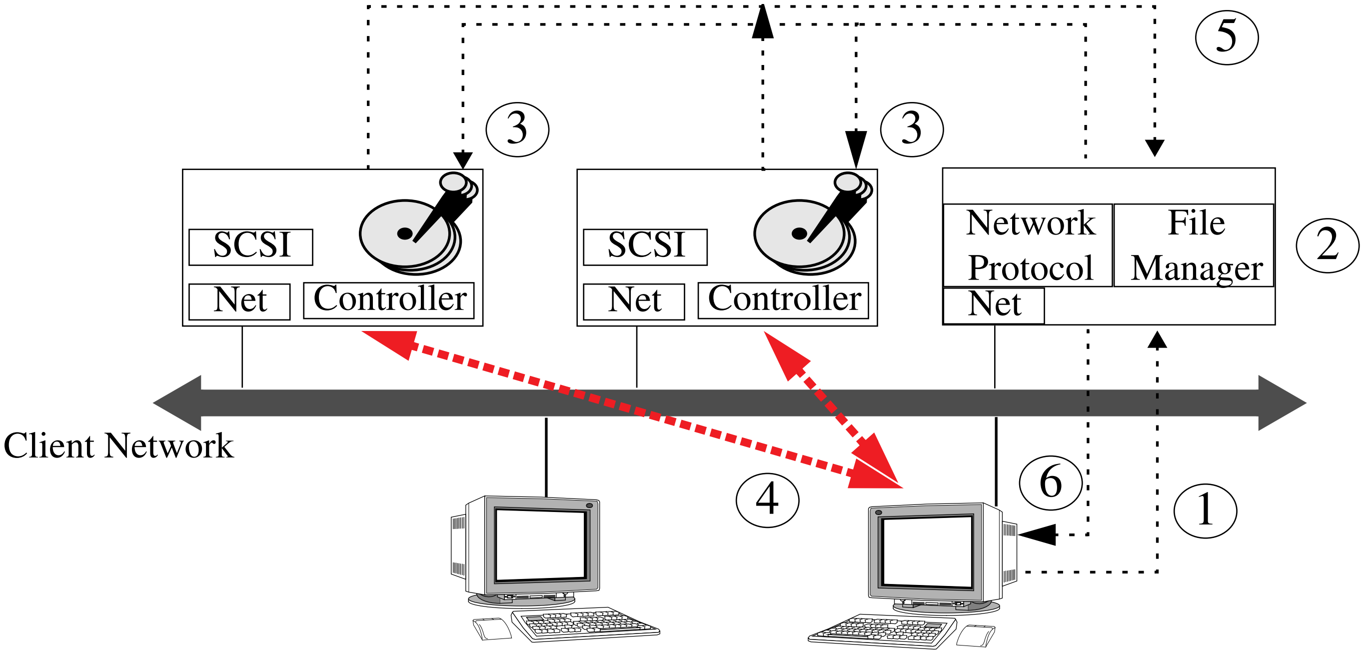
Also Siemens TriCore

Networked SCSI (NetSCSI)

Minimize change in drive HW, SW, IF: RAID-II

- server translates (2) and forwards (3) request (1)
- drive delivers data directly to client (4)
- drive status to server (5), server status to client (6)

Scalable bandwidth through network striping

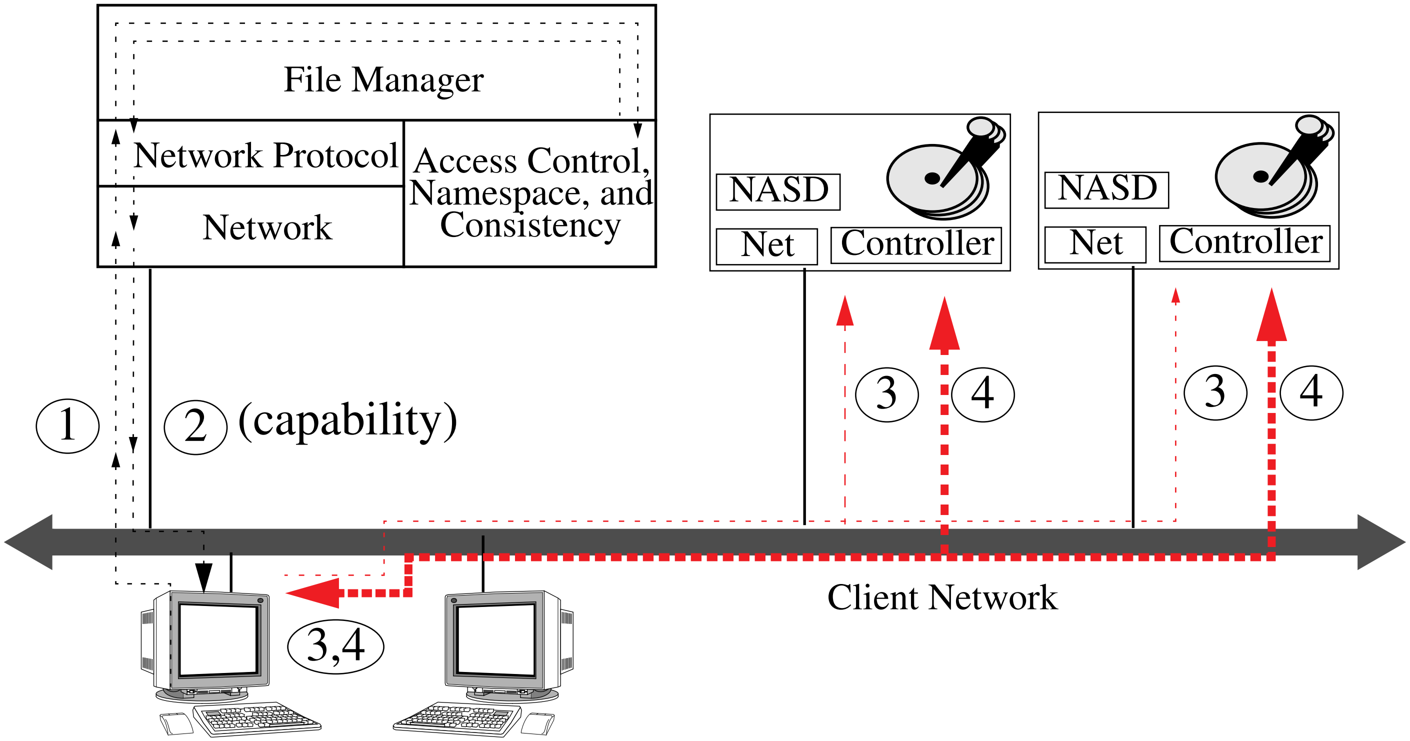


Object Oriented Disk (CMU NASD)

Avoid file manager unless policy decision needed

- access control once (1,2) for all accesses (3,4) to drive object
- spread access computation over all drives under manager

Scalable BW, off-load manager, “file” knowledge



Contrasting Storage Architectures

Server-Attached, Server-Integrated Disk (SAD, SID)

- (specialized) workstation running file server code
- > 35% overhead cost for bandwidth
- striping over servers requires server for servers

Networked SCSI

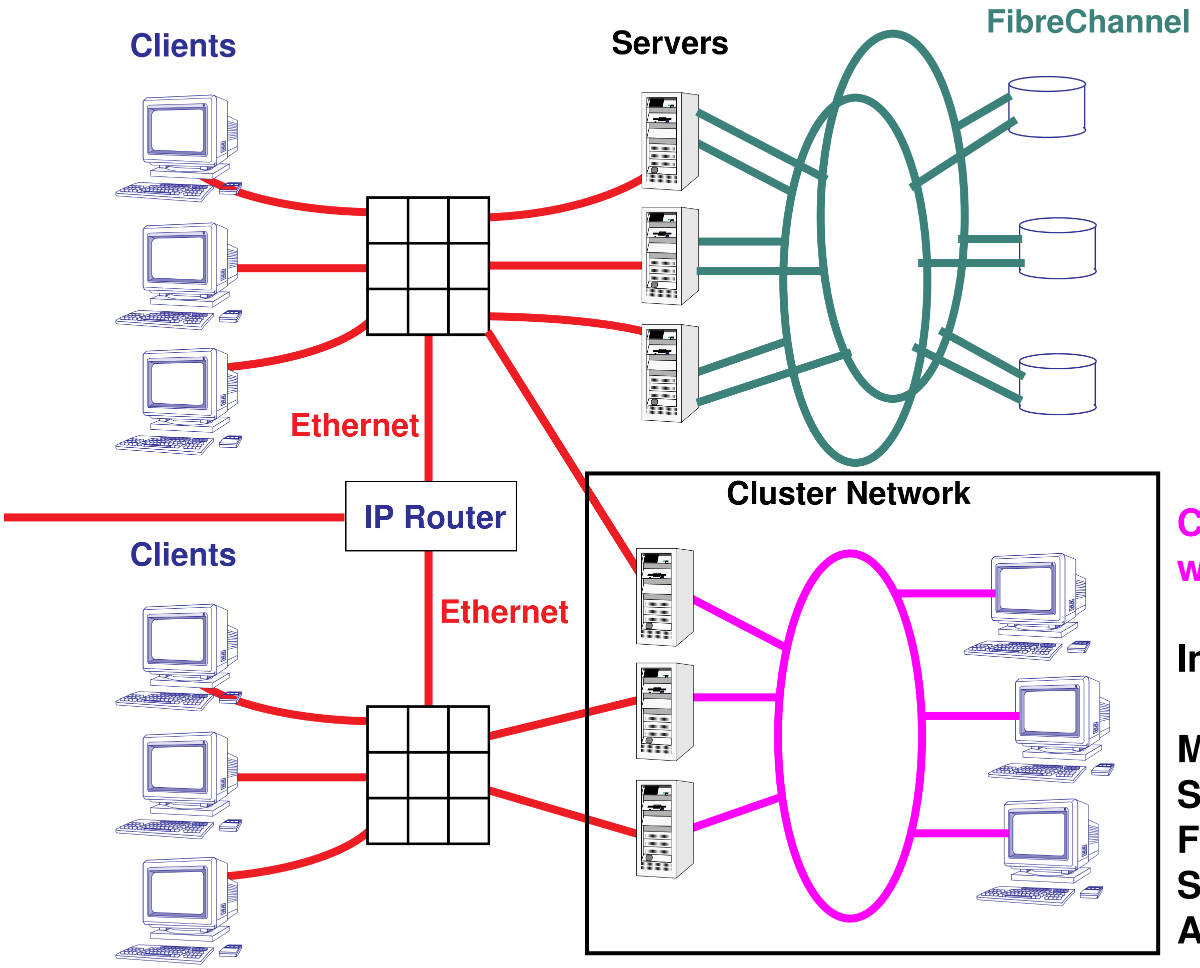
- minimal differences from SCSI; manager inspects requests
- scales massive transfer bandwidth well

Object Oriented Disks

- new (SCSI-4) interface enables direct, preauthorized access
- scales massive, large and small transfer bandwidth well

Emerging commodity cluster nets add new angle

For cost-effective scalable servers



COTS I-connect products with standard API?

Intel/MS/Compaq VI API

Myrinet
Servernet
Fibrechannel?
SCI
ATM

A Wire-Once Vision of Networking

Cluster network is LAN & peripheral interconnect

WAN protocols not used for intra-LAN traffic