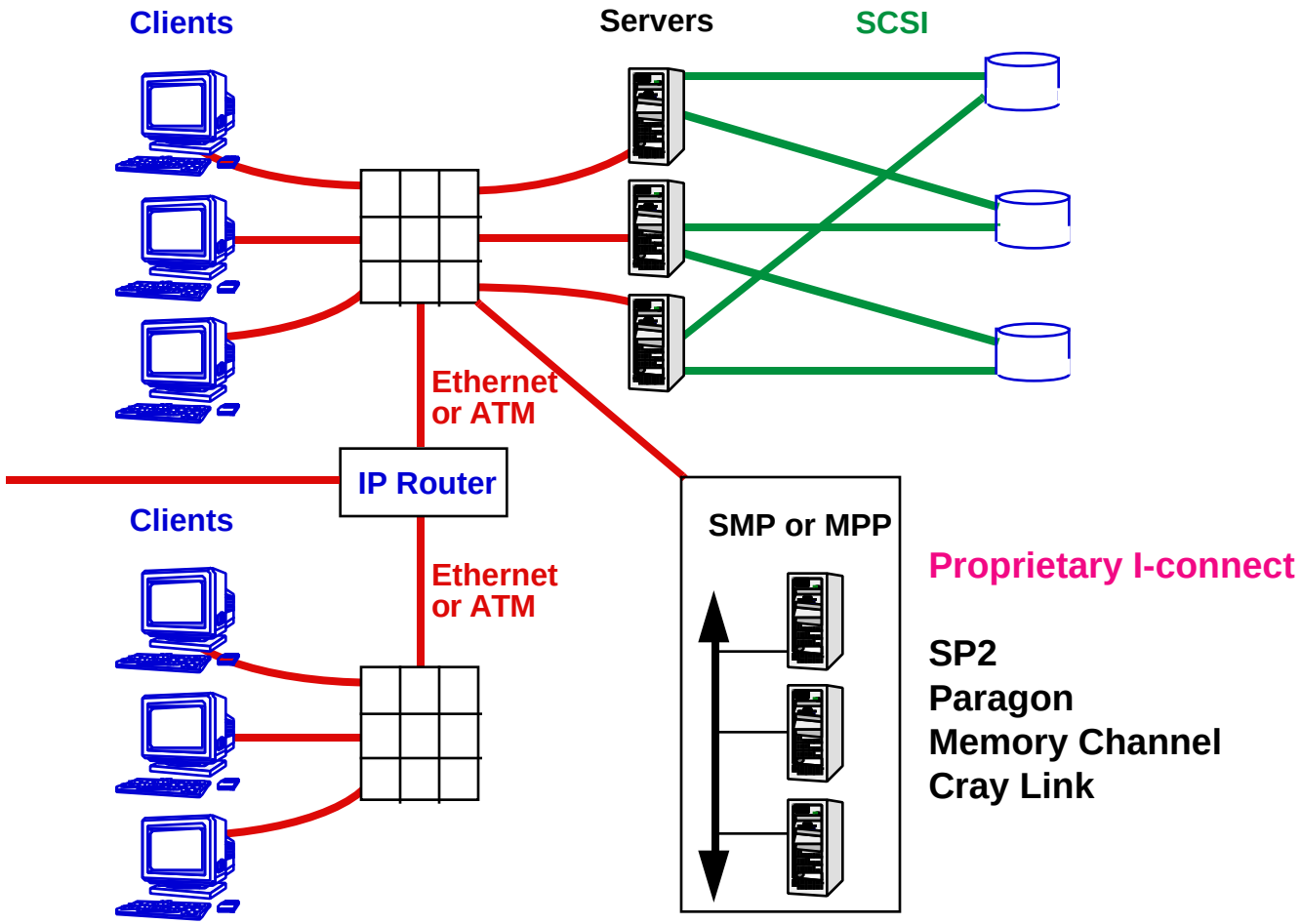
# NSIC/NASD Introduction: Network-Attached Storage CMU's Perspective

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

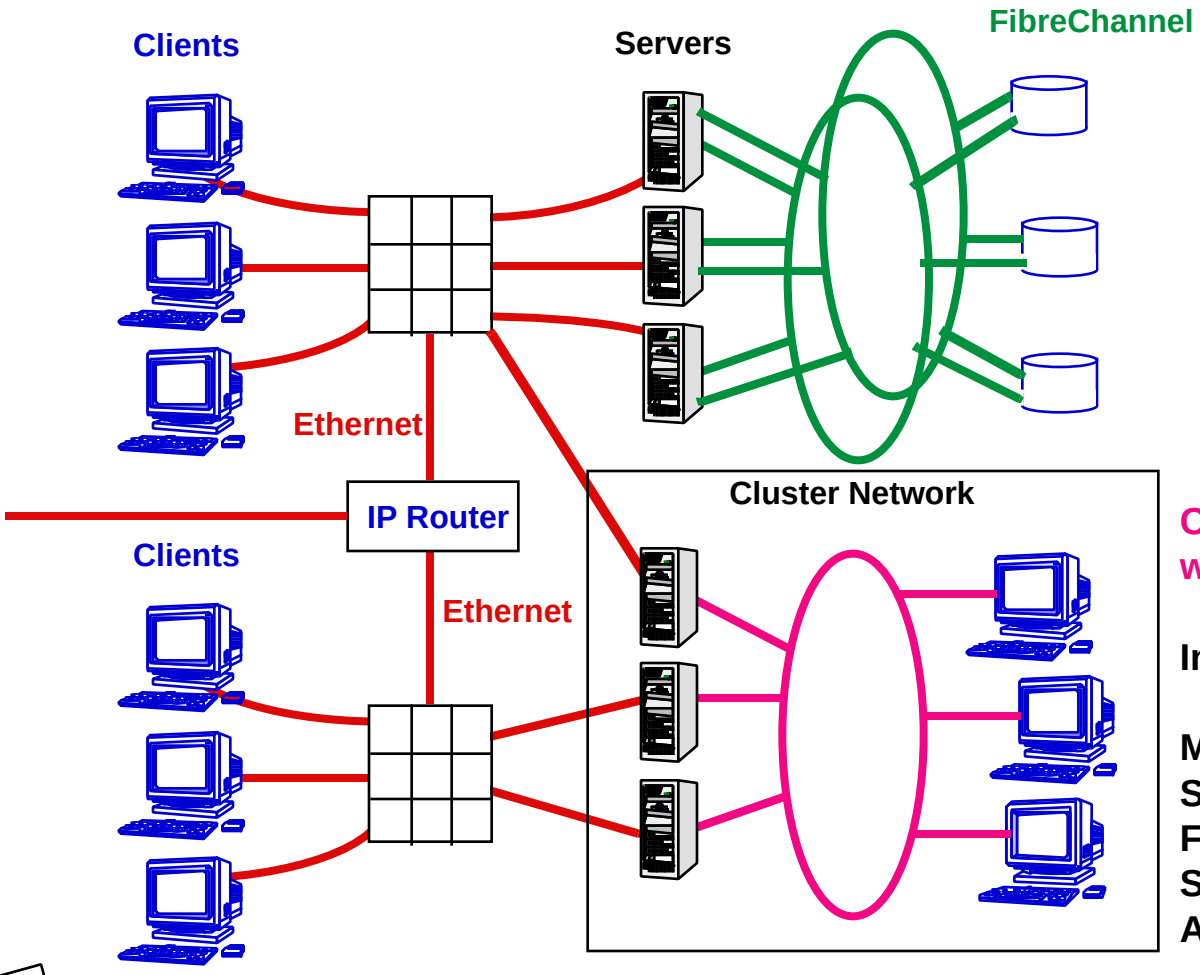
# Endpoint networking world as it is today

## Scalable nets give scalable aggregate BW in LAN



# Commodity peripheral and cluster nets emerge

## For cost-effective scalable servers




**COTS I-connect products with standard API?**

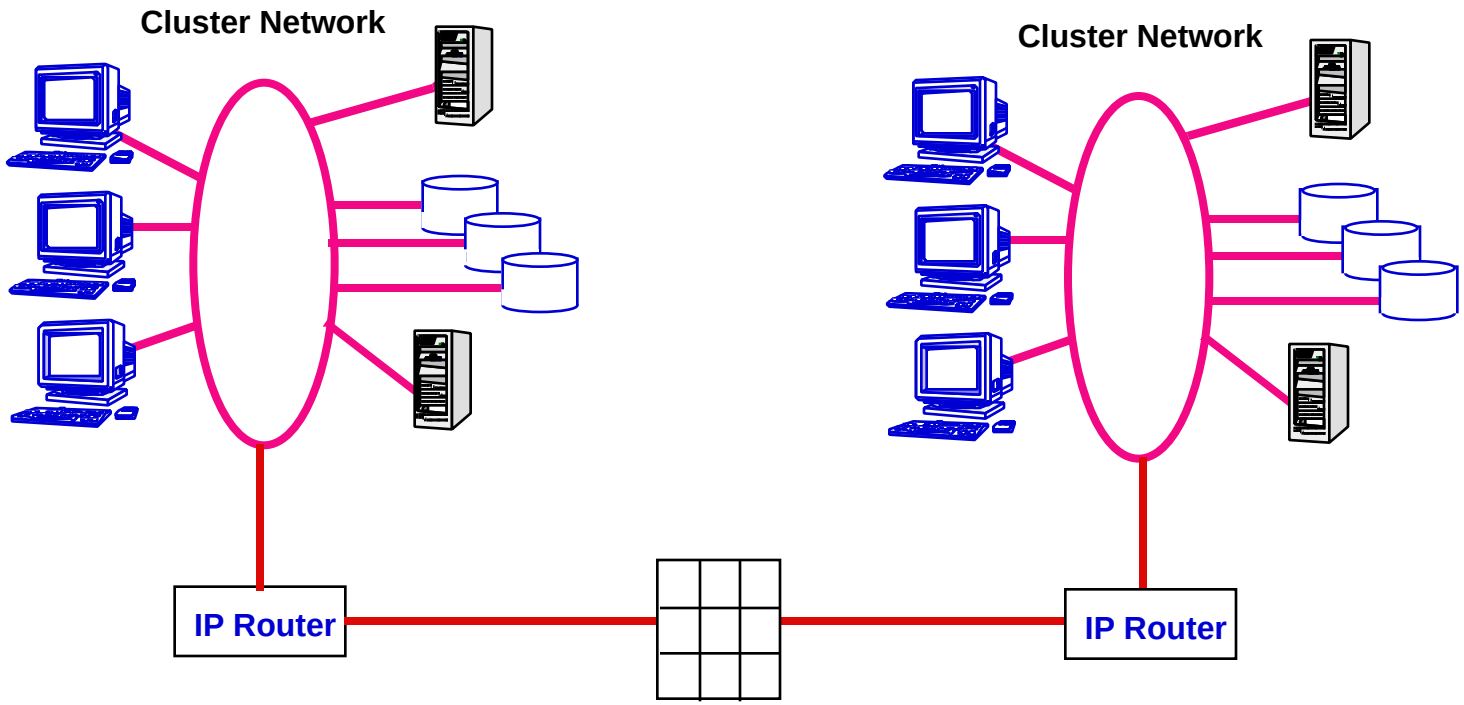
**Intel/MS/Compaq VI API**

**Myrinet**
**Servernet**
**Fibrechannel**
**SCI**
**AN2 (ATM w/ flow control)**

# A Wire-Once Vision of Networking

**Cluster network is LAN & peripheral interconnect**

**WAN protocols not used for intra-LAN traffic**



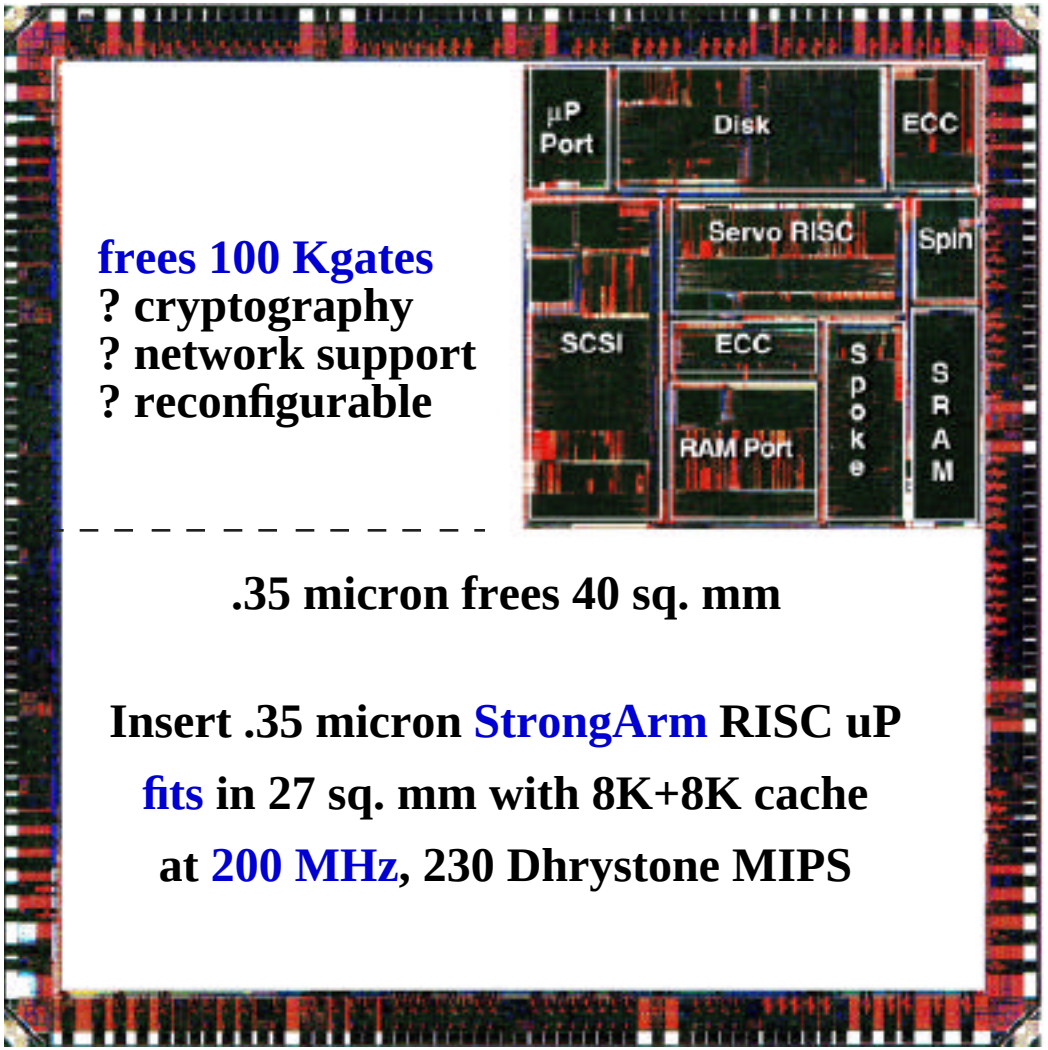
# Are Device Cycles Really Available?

## Quantum Trident drive

- Control: M68020
- Datapath ASIC
- .68 micron in 1997
- 4 indep clock domains, each 40 MHz
  - 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 ASIC++

Current .68 micron chip is 74 sq. mm

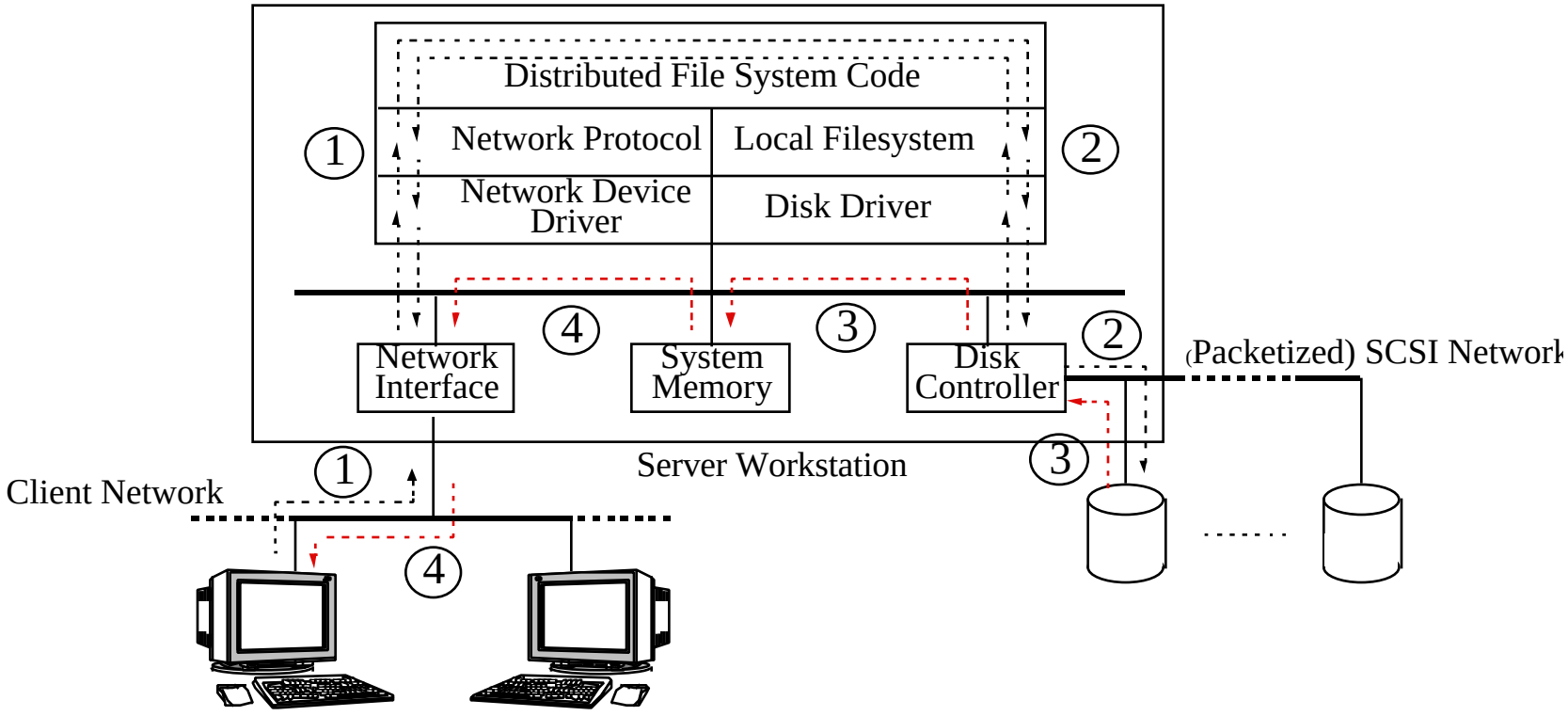


# Problems with current Server-Attached Disk (SAD)

## Store-and-forward data copying thru server machine

- translate and forward request, store and forward data

## Limited bandwidth, slots in low-cost server machine



# The Fix: Partition traditional distributed file server

**Enable direct transfer between client & storage device**

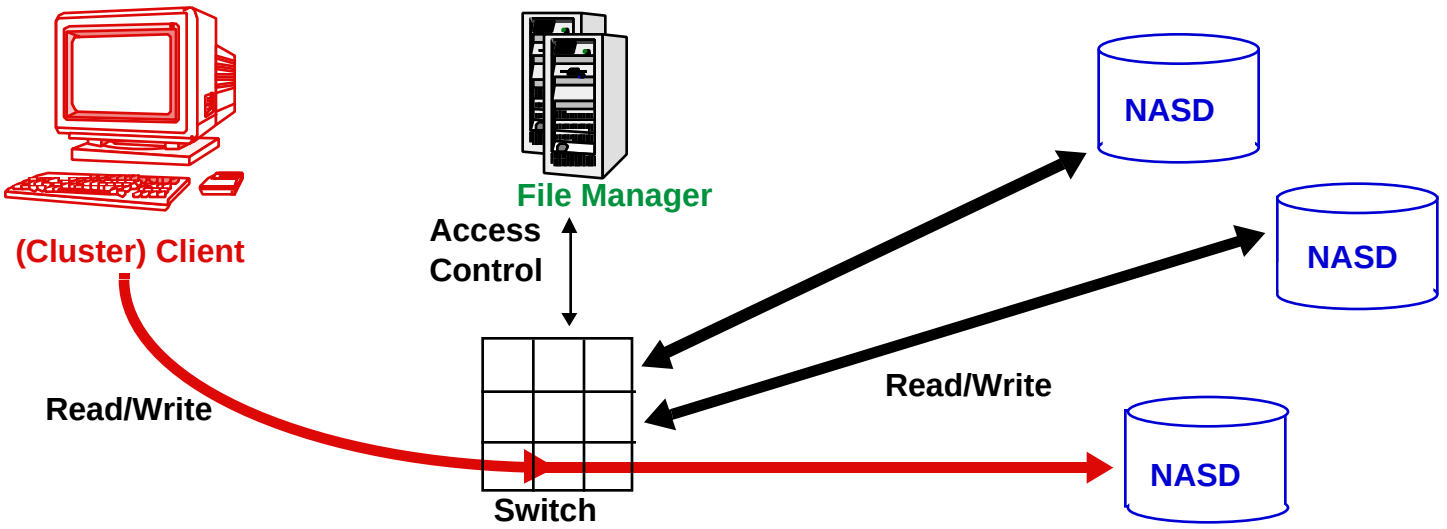
**Low-level networked storage device**

- **direct read/write, high bandwidth transfer**

**Policy moved to file manager**

- **naming, access control, consistency, atomicity**

**Goal of NASD projects – develop the "right" interface**



# Storage industry is ready and willing

## Disk bandwidth: now 10+ MB/s; soon 30 MB/s

- Disk-embedded, high-speed, packetized SCSI
- Eg. 100+ MB/s Fibrechannel peripheral interconnect

## Disk areal density: now 1+ Gbpsi; growing 60%/yr

- Increasing TPI demands more complex servo algorithms
- Put faster RISC processor in integrated function ASIC

## Profit-tight marketplace exploits cycles to compete

- Geometry-sensitive disk scheduling, readahead/writebehind
- RAID support to off-load parity update computation
- Dynamic mapping for transparent optimizations
- Cost of managing storage per year 3-7X storage cost

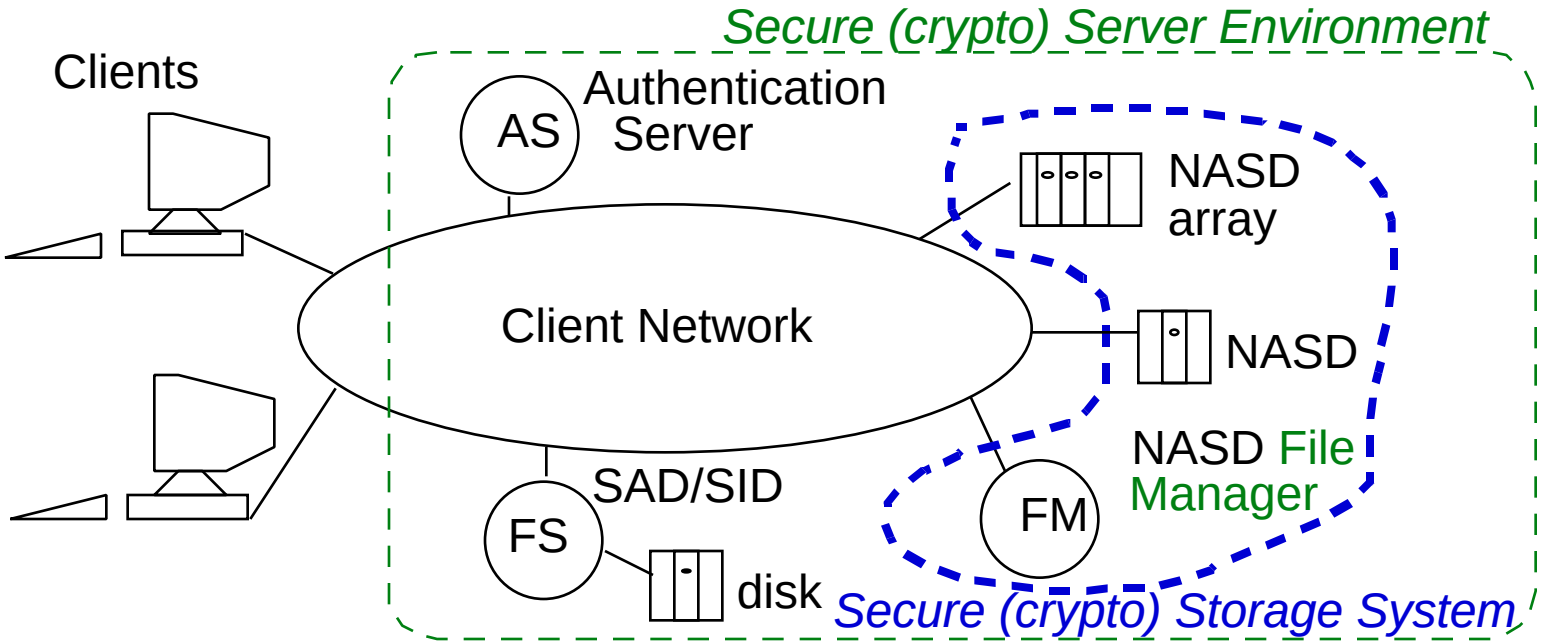
# Security implications of network-attached storage

**Not tied to any specific higher-level security system**

- **ie., not Kerberos, authenticated RPC, x.509**

**Authenticates command to be file manager approved**

- **rests on secrecy of file manager key (hierarchy) only**
- **client's key is computable by drive from file manager's**
- **revocation, restricted operations, auditing supported**

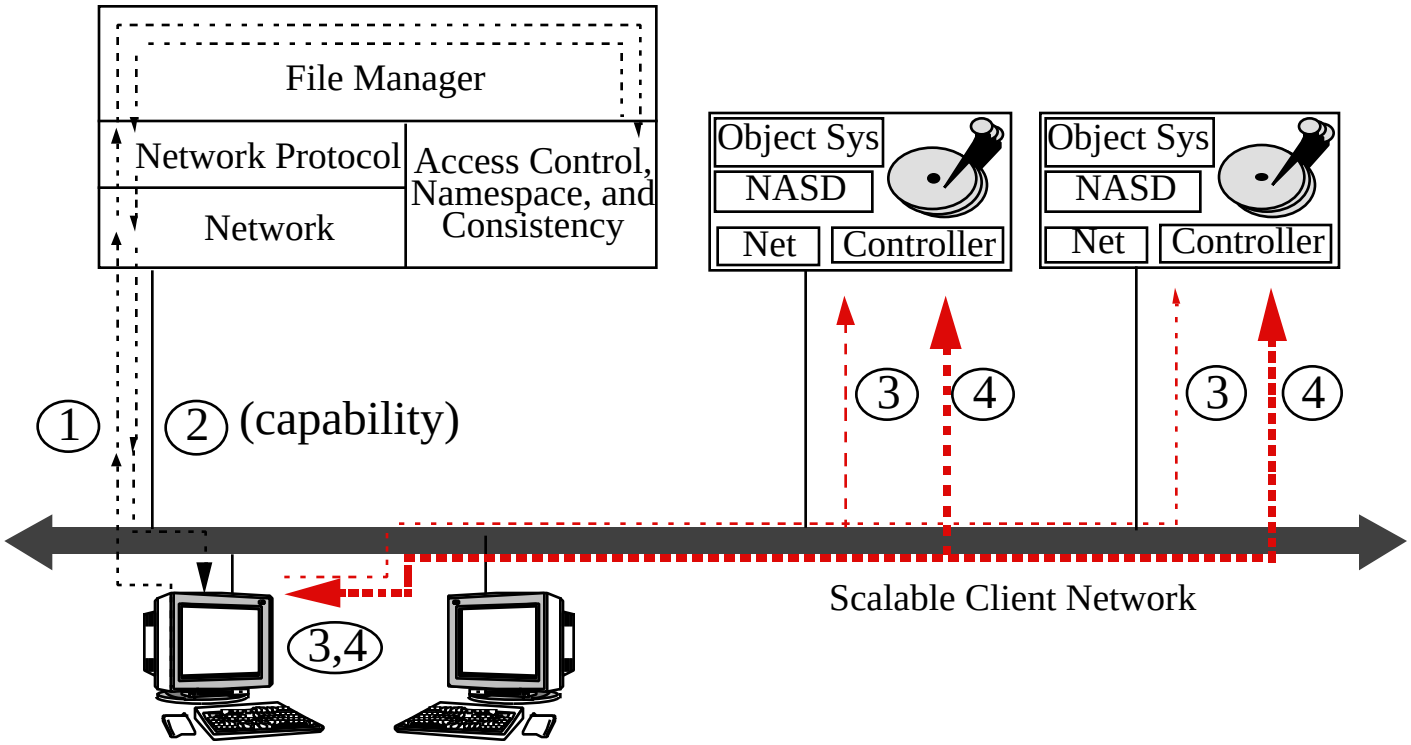


# CMU NASD Prototype Design

## Enable exploitation of all resources in each drive

- network interfaces in proportion to capacity
- storage computation in proportion to capacity

## Scalable client BW, off-load manager, fewer messages

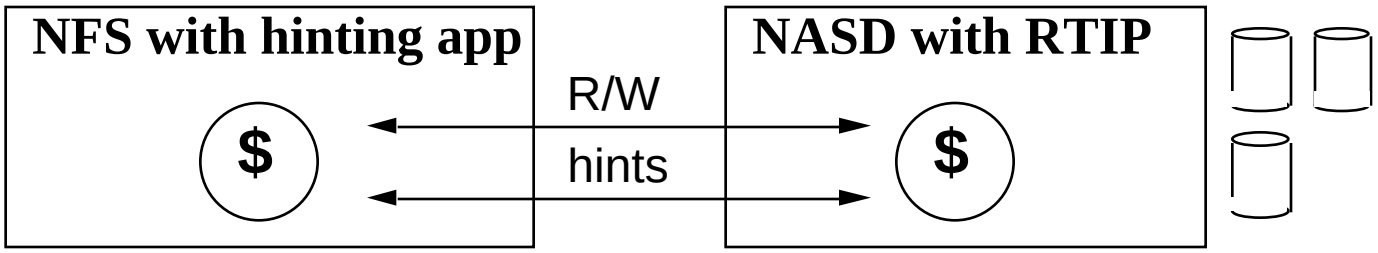


# NASD object interface: Where is file metadata?

**Not at client: don't rest integrity on trusted client**

**Data Layout in storage device ?**

- **avoid distributing per-drive block lists to drive**
- **enables on-drive, drive-subsystem optimization**
  **ie. AutoRAID; deleted space recovery**
  **ie. interposed/stackable NASD - striping, RAID**
  **ie. remote Transparent Informed Prefetching**




- **RTIP in NASD**
- **XDS rendering 25 planes from 64 MB**
- **data striped on 3 disks**

| | NFS |
|---|---|
| Hints | 68s |
| Nohints | 120s |

# Industry NASD collaboration forum

## National Storage Industry Consortium (NSIC)

- launched NASD project April 96
  http://www.hpl.hp.com/SSP/NASD
- mission is to do pre-standards roadmap research for network-attached storage devices and systems
- core group: CMU, HP, IBM, STK, Seagate, Quantum
- quarterly public workshop in addition to core meeting

## Recently ~20 workshop talks

- speakers from HP, STK, Seagate, DEC, Tandem, IDC, CMU, Arizona, MIT, LLNL, USC/ISI, Intel, 3Com, GigaNet,
- attendees also from NIST, IBM, Clariion, Symbios, Compaq, Quantum, EMC, Novell

## Unrelated to Storage Networking Industry Assoc.

- considering appropriate collaboration

# Summary: moving function to storage is multi-win

**Network-stripe storage for scalable bandwidth**

**Drive computational power rapidly growing**

**Industry needs to evolve peripheral network**

**NASD: offload transfer, simple command processing**

**NASD crypto protocol verifies file manager decisions**

**Ported NFS, AFS file managers support more clients**

**Object interface for drive extension, performance opt.**

**Industry group working on standards recommendation**