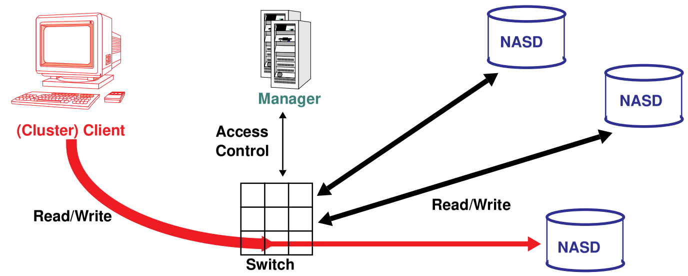
Network-Attached Secure Disks (NASD)

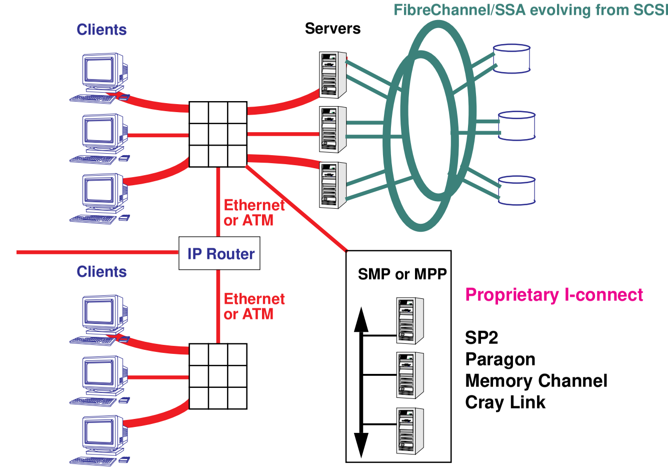
Garth Gibson
http://www.pdl.cs.cmu.edu/NASD

Meet scaling compute needs with storage striped over scalable client network



Endpoint networking world

Scalable nets give scalable aggregate BW internally

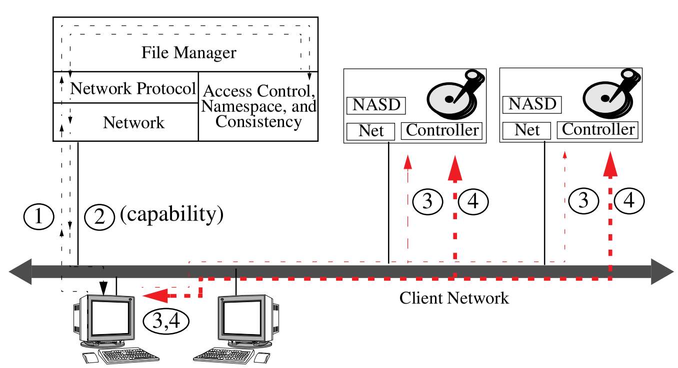


Network-Attached Secure Disk (NASD)

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, fewer messages



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

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

Profit-tight marketplace exploits cycles to compete

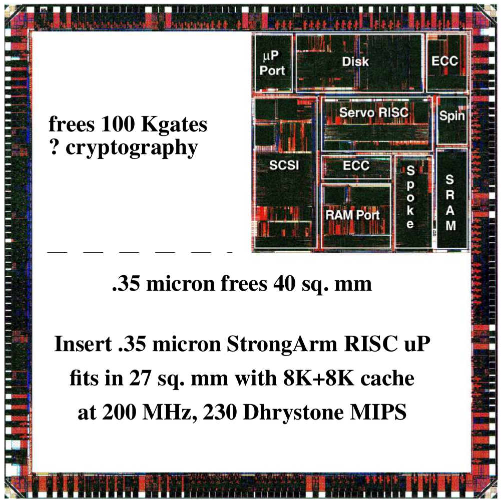
- **Geometry-sensitive disk scheduling, readahead/writebehind**
- **RAID support to off-load parity update computation**
- **Dynamic mapping for transparent optimizations**
- **Cost of managing storage per year 3-7X storage cost**

Large increases in drive MIPS cost-effective

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

Current .68 micron chip is 74 sq. mm



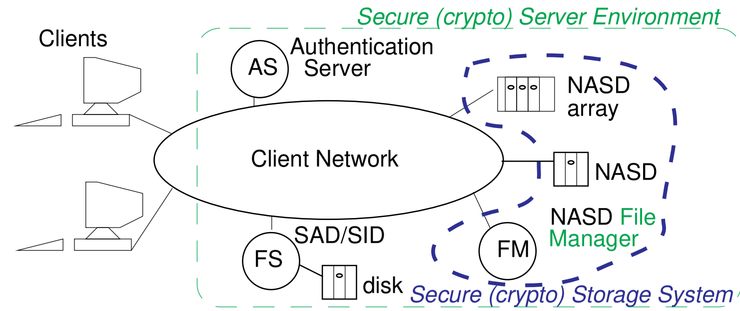
Security implications of network-attached storage

SCSI storage trusts all well-formed commands !

Storage integrity critical to information assets

Firewall is bottleneck, costly, ineffective

Use cryptography same way as currently using ECC



NASD security protocol: integrity protection

Clients “carries” access rights to NASD drive

- manager builds Capability, sends to client to “carry” to drive
- Capability = Digest(Key,Drive,Object,Version,Rights,Expiry)
- Key is secret between manager and drive (really 1 of 4 keys)
- request for Operation on Object sent by client to Drive:
  Operation,Object,Rights,Expiry,Digest(Capability,Operation)

Drive must enforce prior manager authorization

- drive computes capability, operation digest on each request
- manager revokes Capability by 1) letting it expire, or
  2) advancing Object’s Version on drive
- no explicit message to drive with each client open
- drive can reduce digest costs by caching capabilities

Prototyping NASD: NFS & AFS on NASD

File -> NASD object; Directory -> NASD object

NASD object: private metadata, exposed attributes

- allocation: length, blocks used; times: create, data modify
- FS specific: NFS: owner, group, mode
- FS specific: AFS: above and modify time

Operation disposition

- NFS: to drive: get attribute, read, write
- AFS: to drive: FetchStatus, BulkStatus, FetchData (w/cap),
  StoreData (w/cap)
- AFS: Read w/o cap: GetCap (callback, attributes),
  (GetAttr from drive), FetchData
- AFS: Write w/o cap: GetWCap, StoreData,
  ReturnCap (break callbacks)

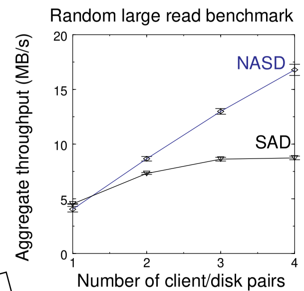
Experimenting with striped NFS-NASD prototype

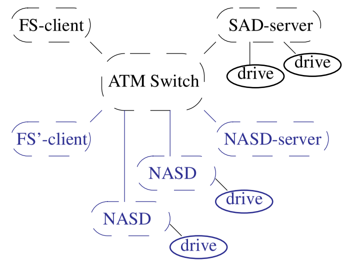
Transparent function extension through NASD stacks

- NFS-NASD FM issues capabilities on a psuedo-object
- Psuedo-object managed by NASD-striper
- After first touch by each, direct client-drive transfers

Experiments on DEC Alpha testbed; DCE on OC3

- user-level NASD client library
- aggregate random large read BW scales with client/drives



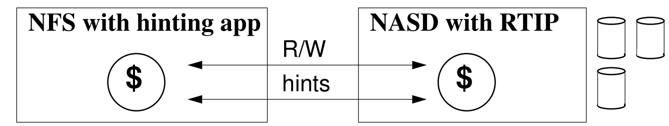


NASD object interface: Where is file metadata?

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

National Storage Industry Consortium (NSIC)

- launched NASD project April 96 (CMU, HP, IBM, STK)
  http://www.hpl.hp.com/SSP/NASD
- signed IP rights sharing agreement Jan 31 97
  CMU, HP, IBM, STK, Seagate, Quantum
- Participants execute independently funded research, sharing issues impacting NASD architecture/interfaces
- quarterly two-day meetings; monthly teleconferences
- host public workshop with each meeting

Recently ~20 workshop talks

- speakers from HP, STK, Seagate, DEC, Tandem, IDC, CMU, Arizona, MIT, LLNL, USC/ISI
- attendees from NIST, IBM, Clarion, Symbios, Compaq, Quantum, EMC, Novell

IDC predicts $11B Net-Attach Storage market in 2000

Related work

Network-attached (secure) storage

- Baracuda, Seagate; DVD, van Meter

Third-party transfer

- RAID-II, Drapeau; PIO, Berdahl; MSSRM, P1244; SCSI

Richer storage interfaces

- Logical Disk, deJonge; Petal, Lee; Attribute Mgd, Wilkes;

Server striping

- Zebra, Hartman; xFS, Dahlin

Capabilities

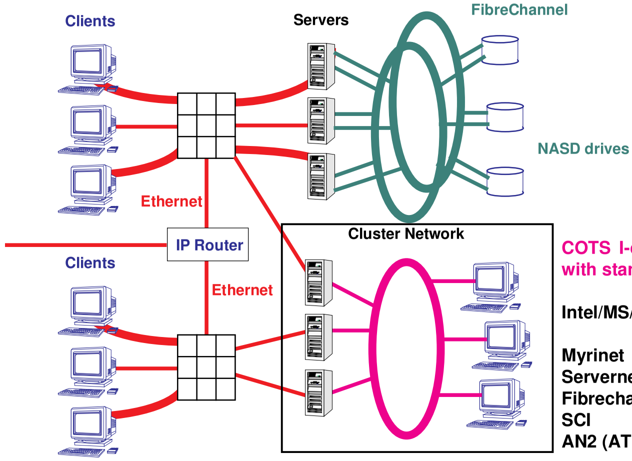
- Dennis66; Hydra, Wulf; ICAP, Gong; Amoeba, Tanenbaum

Application-assisted storage

- Mapped cache, Maeda; Fbufs, Druschel;
  Cooperative caching, Dahlin, Feeley

Critical related work: cluster net API standards

COTS cluster nets 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)

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.

Stacked function layers need not imply copying

Industry group working on standards recommendation