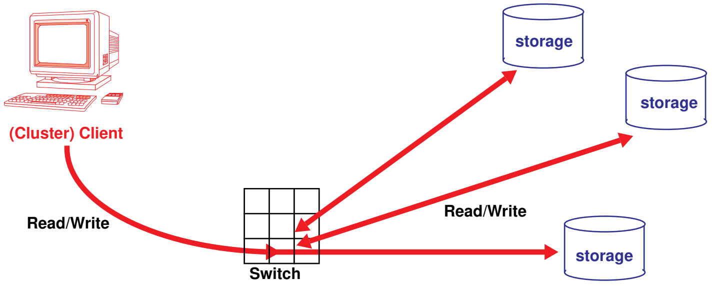
File Server Scaling with Network-Attached Secure Disks (NASD)

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

David Nagle, Khalil Amiri, Fay Chang, Eugene Feinberg, Howard Gobioff,
Chen Lee, Berend Ozceri, Erik Riedel, David Rochberg, Jim Zelenka

Meet scaling compute needs with storage striped over scalable client network



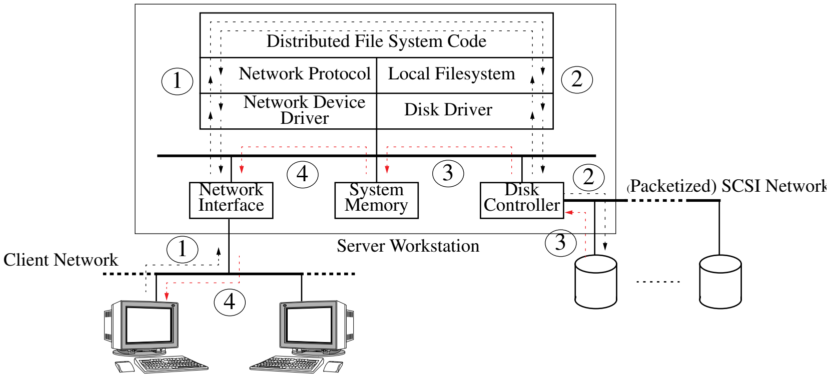
Problems with current Server-Attached Disk (SAD)

Store-and-forward data copying thru server machine

- translate and forward request, store and forward data

Limited bandwidth, slots in low-cost server machine

- server adds > 50% to $/MB and can't deliver bandwidth



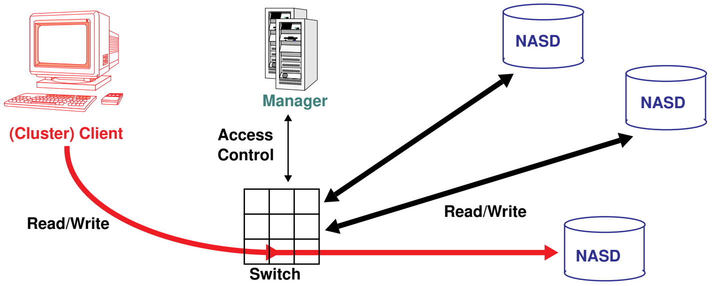
The fix: partition traditional distributed file server

High-level striped file manager

- naming, access control, consistency, atomicity

Low-level networked storage server

- direct read/write, high bandwidth transfer
- function can be integrated into storage device



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
- RISC processor core moving into on-disk 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

NSIC working group on Network-Attached Storage

- Quantum, Seagate, StorageTek, IBM, HP, CMU

What function should be moved?
Taxonomy for Network-Attached Storage (NAS)

Server-Attached, Server-Integrated Disk (SAD, SID)
- (specialized) workstation running file server code

Networked SCSI (NetSCSI)
- minimal differences from SCSI; manager inspects requests

Network-Attached Secure Disk (NASD)
- new (SCSI-4) interface enables direct, preauthorized access

Contrasting extremes: NetSCSI vs. NASD
- both scale bandwidth with large, striped accesses
- what impact on workloads of current LAN file servers?

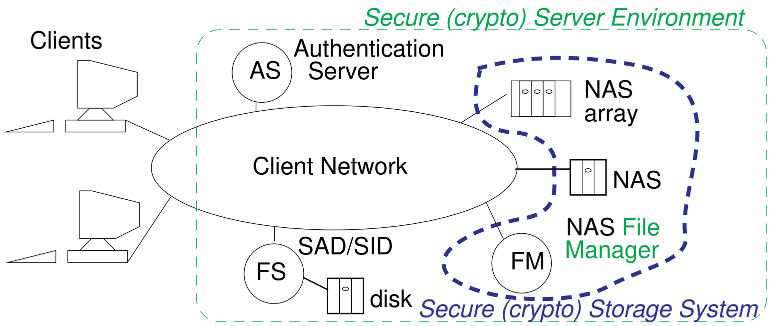
Security implications of network-attached storage

SCSI storage trusts all well-formed commands !

Storage integrity critical to information assets

Firewall is bottleneck, costly, ineffective

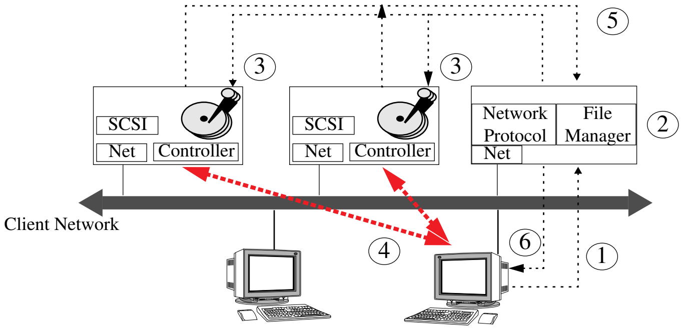
Assuming cryptography used in same way as ECC



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)

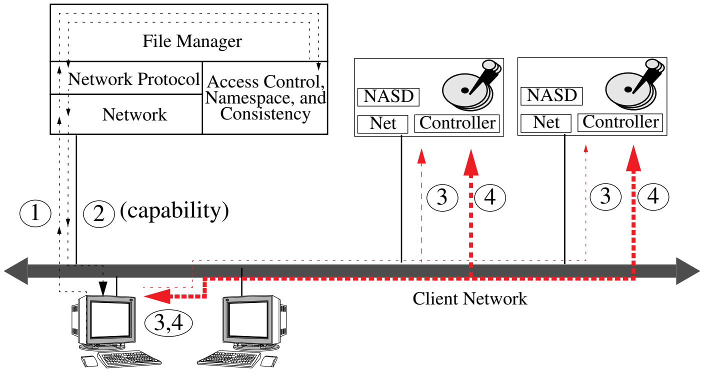
Scalable bandwidth through network striping



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



Impact of NASD vs. NetSCSI on current file systems

Analytic & trace-driven agree; talk presents analytic

Analyze FS traces; instrument SAD server, count instrs

Model change in operation counts and costs at manager

For SAD, use numbers as measured

For NetSCSI, data transfer is off-loaded

- manager does work of 1-byte access per request
- attribute/directory assumed no less work

For NASD, off-load file write and file/attr/dir read

- updates to attributes/directory are no less server work
- manager must do new “authorization” work when file opened (synthesized as first touch after long inactive)

NFS on network-attached storage projections

Berkeley NFS traces [Dahlin94] (230 clients, 6.6M reqs)

Directory/attributes dominate SAD manager work

NetSCSI, therefore, little benefit for manager load

NASD off-loads over 90% of manager load

NFS Operation	Count in top 2% by work (thousd)	SAD		NetSCSI		NASD	
		Cycles (billions)	%of SAD	Cycles (billions)	%of SAD	Cycles (billions)	%of SAD
Attr Read	792.7	26.4	11.8%	26.4	11.8%	0.0	0.0%
Attr Write	10.0	0.6	0.3%	0.6	0.3%	0.6	0.3%
Block Read	803.2	70.4	31.6%	26.8	12.0%	0.0	0.0%
Block Write	228.4	43.2	19.4%	7.6	3.4%	0.0	0.0%
Dir Read	1577.2	79.1	35.5%	79.1	35.5%	0.0	0.0%
Dir RW	28.7	2.3	1.0%	2.3	1.0%	2.3	1.0%
Delete Write	7.0	0.9	0.4%	0.9	0.4%	0.9	0.4%
Open	95.2	0.0	0.0%	0.0	0.0%	12.2	5.5%
Total	**3542.4**	**223.1**	**100.0%**	**143.9**	**64.5%**	**16.1**	**7.2%**

AFS on network-attached storage projections

CMU AFS traces (60-250 clients, 1.6 M reqs)

Data transfer dominates SAD

NetSCSI is able to reduce manager load by 30%

NASD is able to reduce manager load by 65%

AFS Operation	Count in top 5% by work (thousand)	SAD		NetSCSI		NASD	
		Cycles (billions)	%of SAD	Cycles (billions)	%of SAD	Cycles (billions)	%of SAD
FetchStatus	770.5	98.6	37.9%	98.6	37.9%	0.0	0.0%
BulkStatus	91.3	36.6	14.1%	36.6	14.1%	0.0	0.0%
StoreStatus	16.2	3.1	1.2%	3.1	1.2%	3.1	1.2%
FetchData	193.7	83.7	32.1%	24.8	9.5%	0.0	0.0%
StoreData	23.1	15.1	5.8%	3.0	1.1%	3.0	1.1%
CreateFile	12.1	3.7	1.4%	3.7	1.4%	3.7	1.4%
Rename	6.4	1.8	0.7%	1.8	0.7%	1.8	0.7%
RemoveFile	14.6	4.8	1.9%	4.8	1.9%	4.8	1.9%
Others	57.3	13.0	5.0%	13.0	5.0%	13.0	5.0%
Open	480.8	0.0	0.0%	0.0	0.0%	61.5	23.6%
Total	**1665.9**	**260.5**	**100.0%**	**189.4**	**72.7%**	**90.9**	**34.9%**

Recap fundamental modelling results

Network-attached storage offloads file manager
increase manager ability to support storage/clients

NetSCSI offloads transfer only
manager capacity up: 1.6x NFS, 1.4x AFS

NASD offloads transfer, common command processing
manager capacity up: 14x NFS, 2.9x AFS

Implication:
lower overhead cost for manager machines

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**

Summary: moving function to storage is multi-win

Network-stripe storage for scalable bandwidth

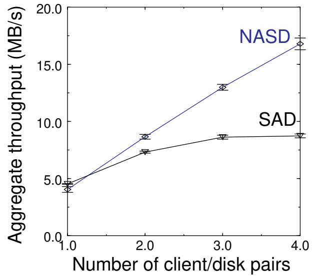
Industry open to evolving functional interface

NetSCSI: server-mgd SCSI; NASD: server-indep access

Both lower manager work; NASD by up to 3-10x

Recent work: prototypes of NASD drives, file systems

Striped NASD/NFS - raw read benchmark



Striped NASD/NFS - file search benchmark