Practical ATA over Ethernet

Ed L. Cashin

SCALE
February 2008

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Topics

The Solution
AoE Characteristics
Linux AoE Support
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced Ethernet
Resources

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

The Solution

Storage needs
External storage
AoE

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Storage needs

- greater storage capacities needed
- greater storage flexibility needed

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

External storage

- External storage allows a greater number of disks to be used.
- Storage outside the box can be used in new configurations.
- ATA over Ethernet provides one mechanism.

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

AoE

ATA over Ethernet (AoE)

- uses familar networking hardware
- requires minimal configuration
- is simple
  - no Internet Protocol
  - dedicated AoE storage networks

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

AoE Characteristics

Protocol basics
ATA basics
AoE details

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Protocol basics

Characteristics of ATA over Ethernet

- “shelf” and “slot” AoE addresses
- ethernet (“MAC”) address
- discovery mechanism
- ATA commands

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

ATA basics

The most important ATA commands are,

- device identify
- read
- write

The read and write commands can use a 48-bit logical block address to indicate where the I/O will be performed on the device.

LBA 48 supports up to 281,474,976,710,655 sectors, or

140737488355327	KiB	144115188075855	KB
137438953471	MiB	144115188075	MB
134217727	GiB	144115188	GB
131071	TiB	144115	TB
127	PiB	144	PB

Table 2.1 LBA 48 maximum

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

AoE details

- more than one command can be sent at a time
- commands can be retransmitted
- AoE target can have multiple ethernet interfaces
- AoE initiator can have multiple ethernet interfaces
- the AoE config string provides locking primitives

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Linux AoE Support

The initiator
The aoetools

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

The initiator

The aoe driver for 2.6.x Linux kernels is an AoE initiator.

- Linus Torvalds' git tree
- Coraid website has backwards compatible aoe driver

New features generally migrate from the Coraid website "standalone" driver to the in-kernel driver after settling into state where a coherent set of changes can be submitted.

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

The aoetools

The aoetools package contains utilities that,

- support the aoe driver
  - aoe-stat displays the targets the driver has discovered
  - aoe-interfaces limits the local ethernet interfaces used for AoE
- perform convenient functions
  - aoeping queries a target from user space
  - aoecfg can read and manipulate the AoE config string on the target

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

The aoetools project at sourceforge has packages besides the aoetools themselves.

- cec — connects to the Coraid Ethernet Console
- srcat — reads data stored on disks that were in an SR
- vblade
- kvblade

The Solution
AoE Charac- ...
Linux AoE ...
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced ...
Resources

AoE Targets

Software targets
Hardware targets

There are hardware and software-based AoE targets.

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Software targets

- vblade - user space minimal target
  - uses raw sockets
  - exports any file, special or regular
  - can be handy
- kvblade - kernel software target
  - faster
  - alpha software
  - needs a maintainer
  - a reimplementation of Lelik P. Korchagin's vblade-kernel
- qaoed - multithreaded target
- aoeserver

The Solution
AoE Charac- ...
Linux AoE ...
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced ...
Resources

Hardware targets

- one target per disk
  - Coraid PATA EtherDrive storage blade
  - Welland Xpress
- one target per aggregation of disks
  - Coraid SR logical blades are based on RAIDs
  - Coraid VS exports storage based on storage extents

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Examples

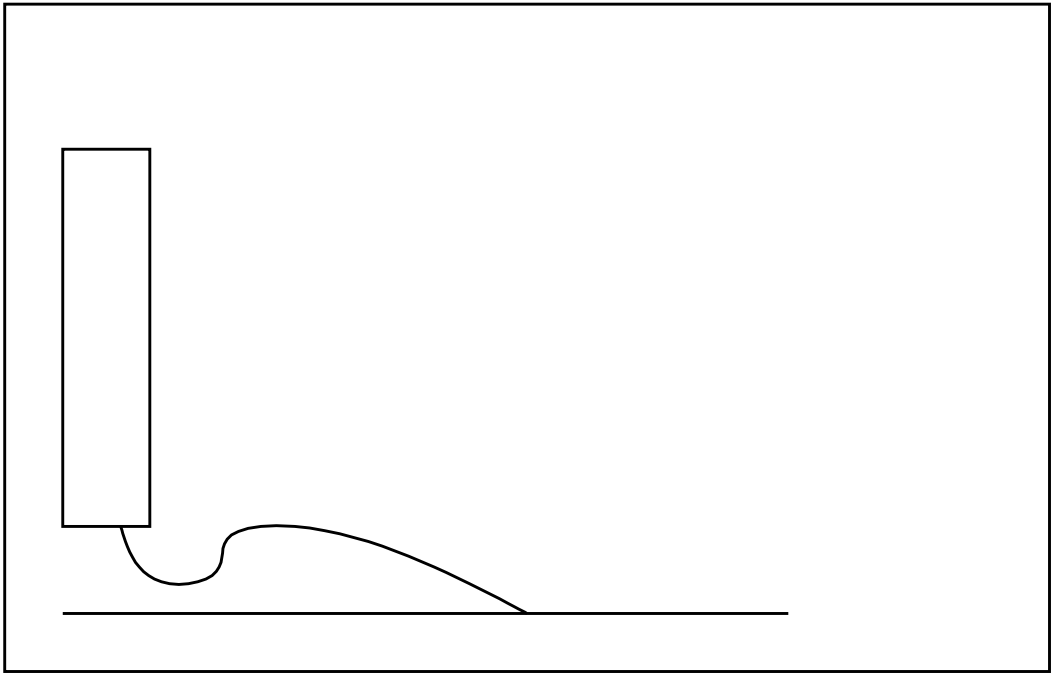

Figure 5.1 Internal Hard Drive

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

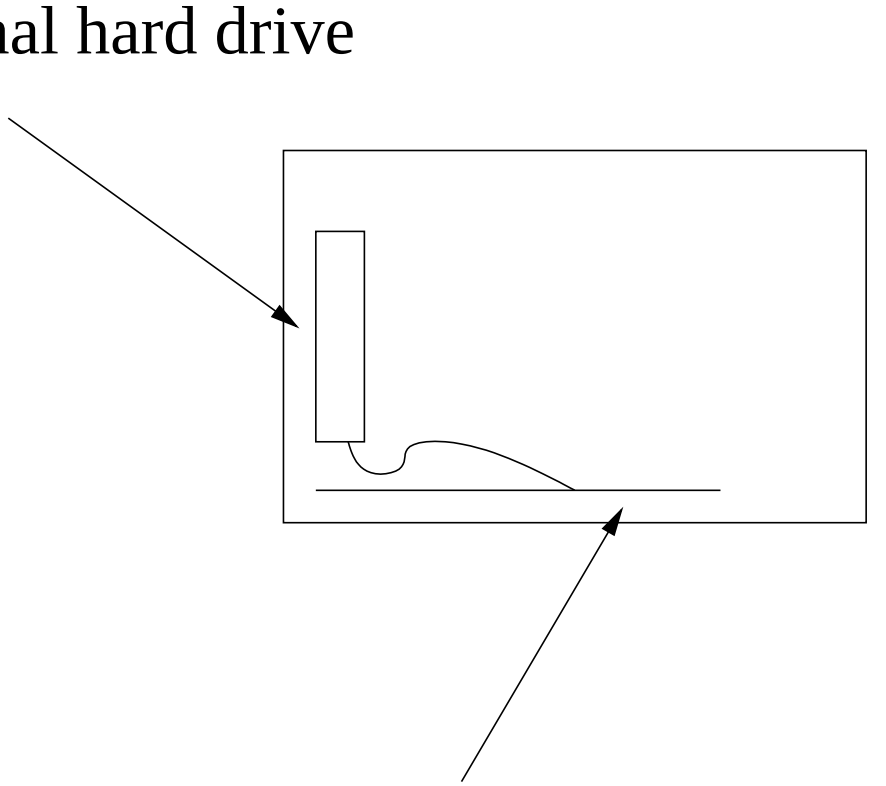

Figure 5.2 Inside the box

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

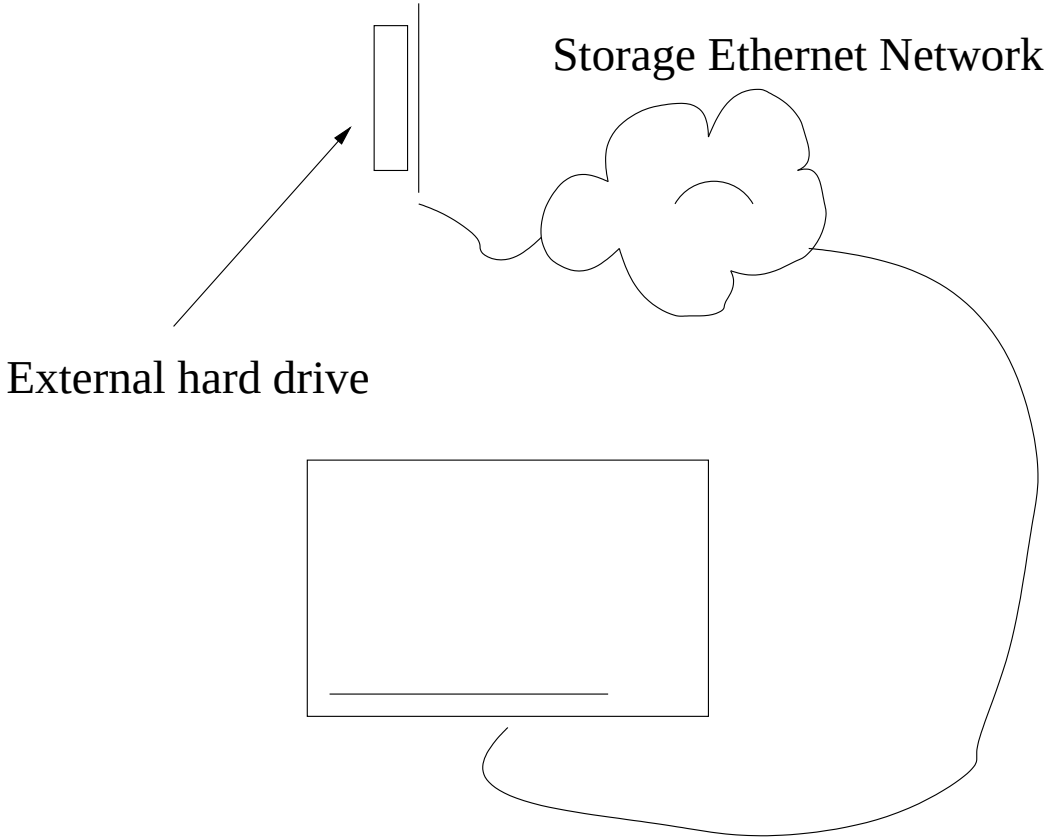

Figure 5.3 Outside the box: EtherDrive storage blade

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

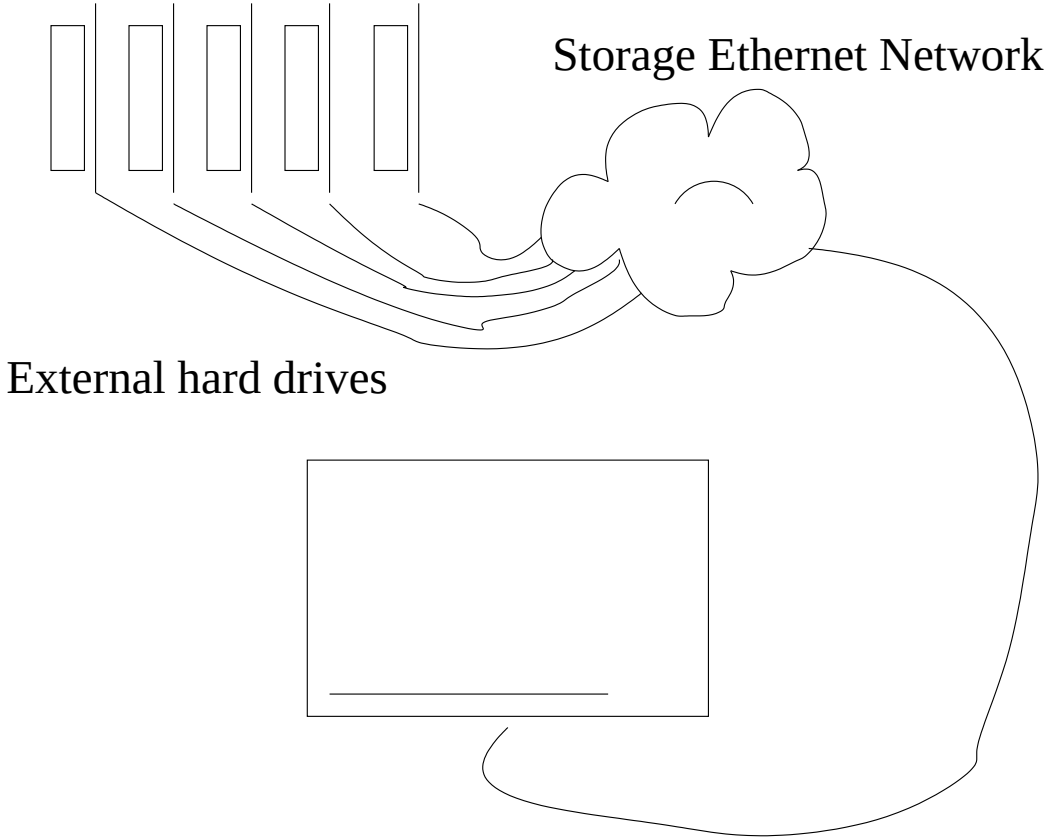

Figure 5.4 EtherDrive storage blades in a shelf

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

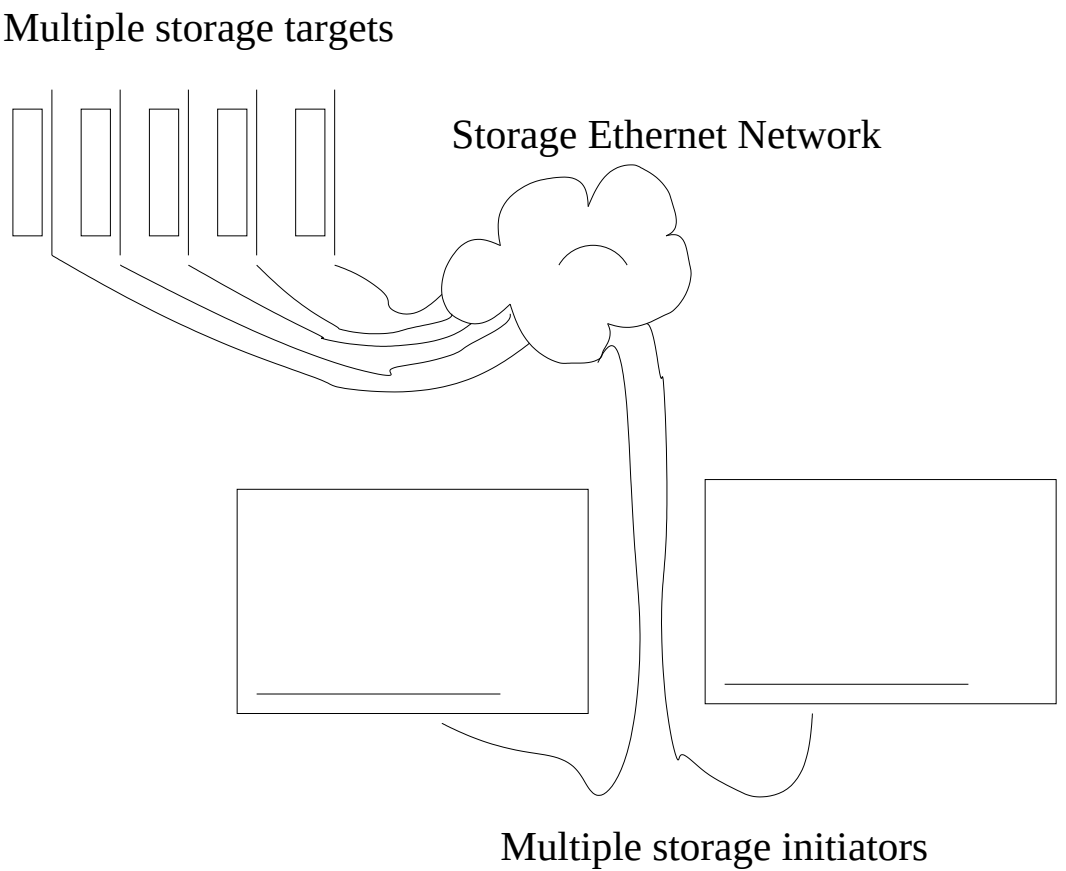

Figure 5.5 Multiple AoE initiators

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

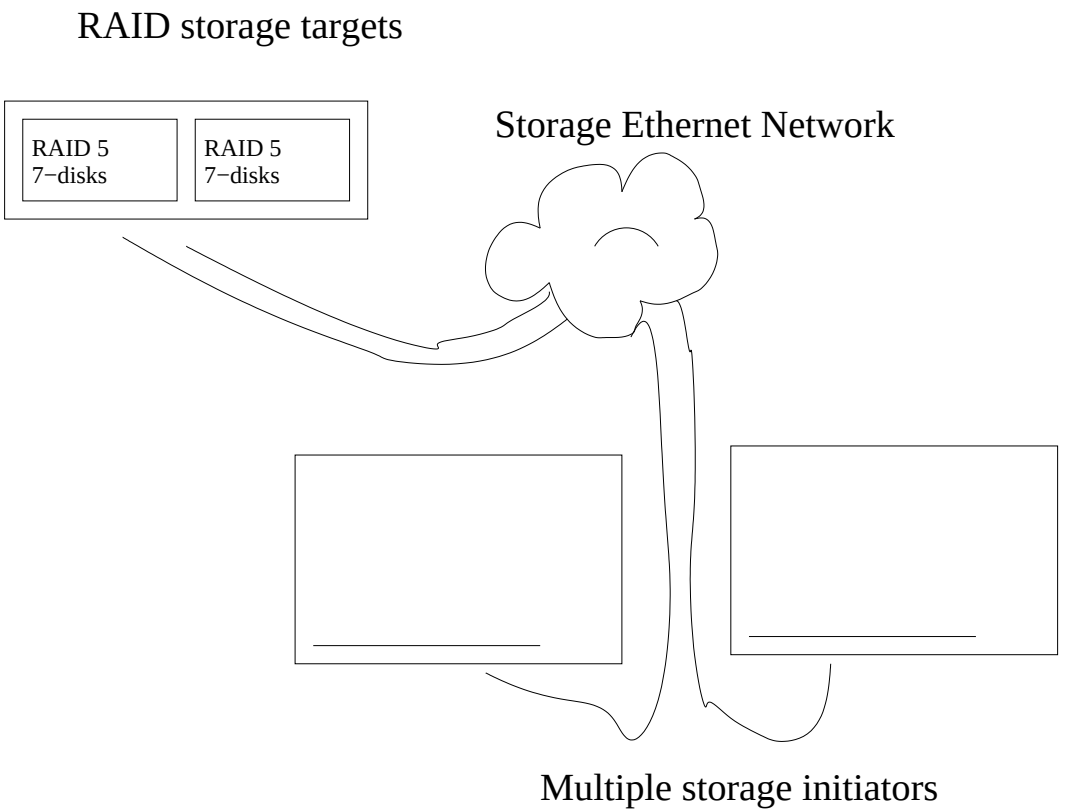

Figure 5.6 "Logical" blades in an SR

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

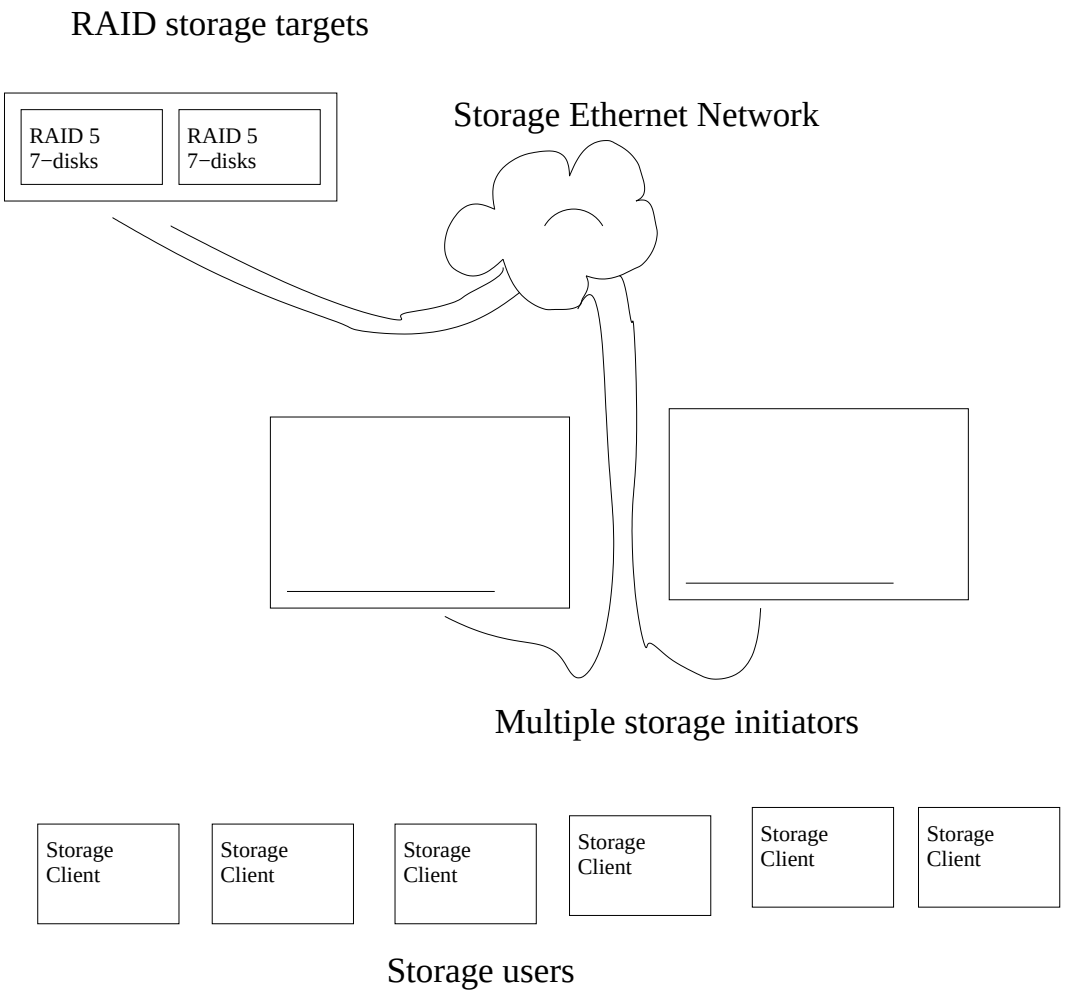

Figure 5.7 Exporting storage from the initiators

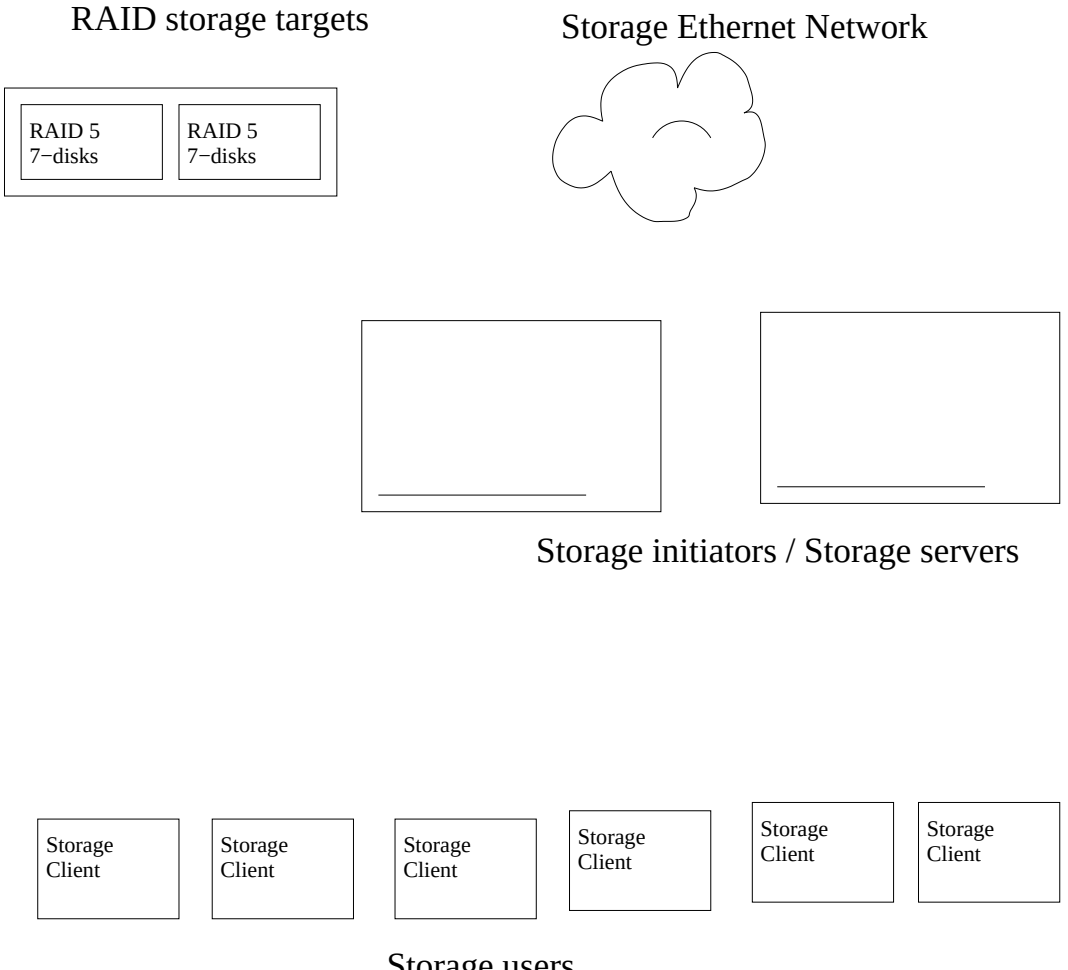

Figure 5.8 Exporting storage to physical clients

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

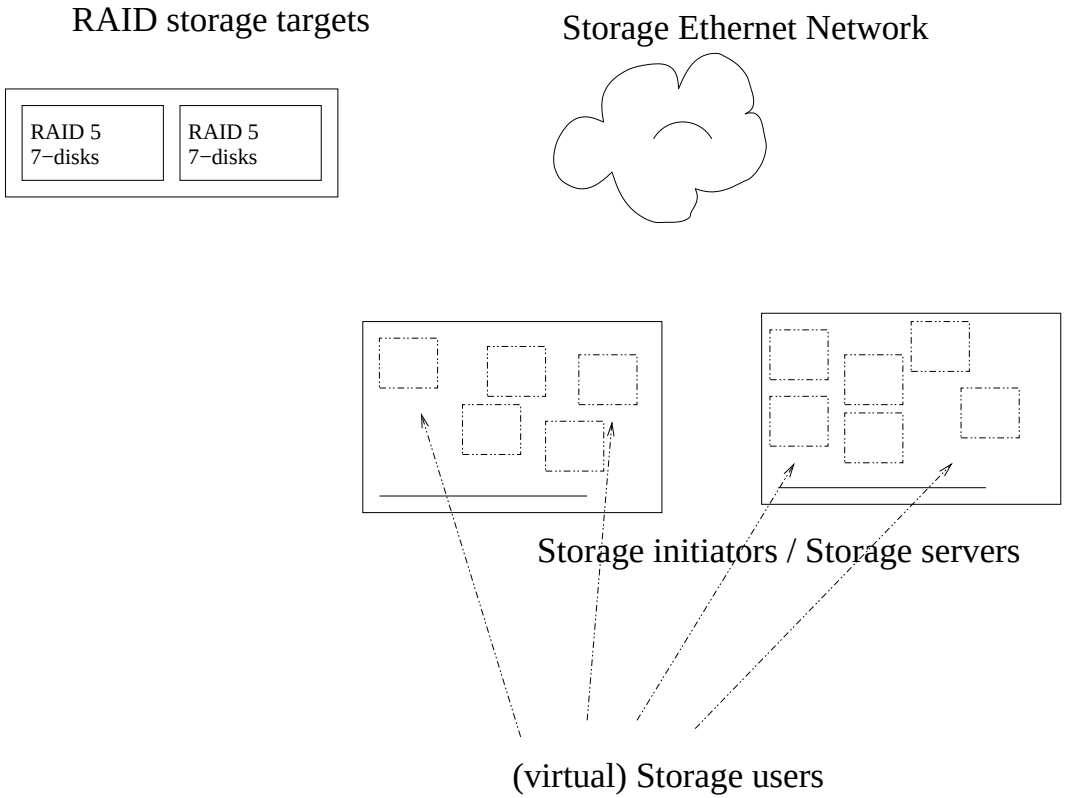

Figure 5.9 Exporting storage to virtual clients

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Filesystems

Single host
Cluster
Distributed

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Single host

Traditionally, it is an error to mount a filesystem that is already mounted.

Some favorite filesystems like ext3 cannot use the large capacities that are possible with external storage.

Selected interesting single-host filesystems for AoE are,

- XFS — from SGI, specializing in large files
- JFS — from IBM, a general purpose journaling filesystem (limited quota support)
- ext4 — the successor to ext3 supports file extents
- Btrfs — from Oracle, “A Checksumming Copy on Write Filesystem”

Some cluster filesystems (GFS, for example) can be used on a single host.

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Cluster

A cluster of AoE initiators can mount a filesystem on shared AoE storage if they cooperate.

Selected cluster filesystems are,

- GFS — the filesystem from RedHat's Cluster Suite
- GPFS — IBM's cluster filesystem

Both physical and virtual hosts can participate in a cluster.

The Solution
AoE Charac- ...
Linux AoE ...
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced ...
Resources

Distributed

- PVFS — from Clemson University
- Ceph — “petascale distributed storage”

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Storage Layers

- storage device, e.g., hard drive
- partitions (MS-DOS, GPT)
- RAID (md, SR)
- storage virtualization (LVM, VS)
- filesystem (single-host, cluster)
- loop (encryption, compression)

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Enhanced Ethernet

Jumbo frames can provide a marked improvement in performance but are not supported by all hardware.

The new 10 gigabit hardware is still more expensive but offers greater throughput.

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources

Resources

- ATA Protocol Specification
  http://www.coraid.com/documents/AoEr10.txt
- The aoetools at SourceForge
  http://aoetools.sourceforge.net/
- The Coraid Linux Support Page
  http://www.coraid.com/support/linux
- http://en.wikipedia.org/wiki/ATA_over_Ethernet

The Solution
AoE Charac- ...
Linux AoE ...
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced ...
Resources

fin

The Solution
AoE Charac- . . .
Linux AoE . . .
AoE Targets
Examples
Filesystems
Storage Layers
Enhanced . . .
Resources