Object Based Storage: A Vision

- Make scalable computing possible
- Make systems more usable, manageable
- Effective, inherent security

“The Storage Utility”
Resources for OBSD appear to be possible, but cost pressure severe
NSIC Network Attached Storage Devices

- IP Agreement signed 2/97
  - Goal: pre-competitive research collaboration
  - Focus: Server Class Storage Management

- Participants
  - HP - John Wilkes, Richard Golding, Lou Schreier
  - IBM - Paul Hodges, Jack Gelb, Greg Van Hise
  - Quantum - Satish Rege, Paul Massiglia, Steve Byan, Geoff Peck
  - STK - Chuck Milligan, Scott Robb, Jim Hughes
  - Seagate - Dave Anderson, Chris Malakapalli
  - CMU - Garth Gibson, Bill Courtright, David Nagle, ...

- Work Product
  - Proposal for an Object Based Storage Device protocol standard
  - Concepts for attribute-based storage management, security model
Activities

- Private quarterly meetings - discuss research, direction
  - First - March 1995
- Public meetings - education for NASD group, public
  - March 98 - general
  - June 98 - Intelligence in drives, Active Discs
  - September 98 - Networking for Storage
  - May 1999 - File Systems
  - August 1999 - Data Bases

Protocol Proposal

- Requirements document posted to NSIC web site June 1998
- Continuously revised
- Will be presented to T10 on November 9th

- Long CDB approved by SCSI
Research Objectives: Find Solutions

Scalable Computing
- Shared access to data
- Heterogeneous computing
- Dynamic scaling without interruption

Storage Management
- Today, more expensive than storage itself
- Platform independent capability
- Goal is self managed storage
  Scales with storage
  Managed by policies and attributes
OBSD: Long CDB

- Proposed on 11/2
  - >80 bytes realistically used
  - 256 total CDB byte limit
  - Bytes 0-9 never encrypted
- Encryption ID: Covers bytes 8 - n
  (Encrypted or not)
- Action code: NAS command
  Determines values for Bytes 11 - n

Now part of SCSI Standard!

<table>
<thead>
<tr>
<th>Byte</th>
<th>7</th>
<th>6</th>
<th>5</th>
<th>4</th>
<th>3</th>
<th>2</th>
<th>1</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Long Command Op Code (7Fh)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>Control byte</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Encryption Identification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Reserved</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Additional CDB Length (n-7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 (MSB)</td>
<td>Action Code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9 (LSB)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Action code specific fields</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Command Set:

- **Object specific**
  - Create Object
  - Open Object
  - Read Object
  - Write Object
  - Append Object
  - Close Object
  - Remove Object
  - Import Object
  - Flush Object

- **Group**
  - Create Object Group
  - Remove Object Group

- **Other**
  - Set Attribute
  - Get Attribute
  - Flush
  - Format OBSD
OBSD Some Key Issues/Features

- Byte addressing only
- Partitions = Object Groups
  - None - majority position
  - Needed for legacy support
  - Needed, and lots = 4 byte address
- Open & Close - polar division
  - Needed for QoS specification
  - Needed for management support
  - Waste of time
- Object ID
  - Set by OBSD or Requester
  - Length - 32, 64, 128 ??
- Sessions

Partitions
- Not divisions of capacity
- Sets of Objects
- Can have capacity quota

Open/Close
- Frame sessions
- Sets QoS parms
- Not equated to file system Open ...
OBSD: What About Aggregation

- What is an OBSD: A New Boundary
  - Not just a disc drive
  - Tape drive, Tape library, Disc Array, Jukebox, etc

- Aggregation - objects across OBSD’s
  - Three kinds
    - For redundancy: RAID, mirroring
    - For Performance: Striping
    - For Capacity: Spanning
  - Two solutions
    - Disc Arrays
    - Do on an Object basis
  - Possible also with the Object abstraction
    - Mapping function required
    - RAID 5 harder, but doable
Data stored as objects (files), not sectors
- Still supports OS’s file structures
- Disc does space management
- Disc knows when a file is open, in use, ...

OBSD Advantages
- OS workload greatly reduced
- Space management scales with # of drives
- Storage can be self-managed
- Storage can do off-line work
  - Copy, Mirror, Backup, etc.

Red line = sector request
Blue line = file/object request
Why Objects

- Key to storage participating in management
- Enables storage supported Quality-of-Service agreements
- Heterogeneous computing made easier(?)
- Performance opportunities
  - Metadata never leaves device
  - More effective cache with Open/Close visibility
  - More knowledgeable prefetching
  - Object level LOCKS + VIA => direct application (DBMS) access
OBSD + VIA = Fast Path to Data

- VIA: Defined by Microsoft, Intel, Compaq
  - Provides shortcut to wire
  - Very low latency communications

- OBSD + VIA
  - Open Objects through OS
  - Then application goes right to disc
  - “Saves 25% of CPU”
Each level was in the host at some point!
Each advancement was met with resistance
Eventually advantages of new interface became compelling
OBSD and SNIA and ANSI

- Object Based Storage affects more than devices
  - Need participation, input from broader industry
    - File system
    - Middle ware suppliers
    - Data based developers
    - Hardware vendors
- Need collaboration with ANSI SCSI committee
  - SNIA not a standards body
  - ANSI lacks broad representation
OBSD and SNIA and ANSI

- NSIC members committed to SNIA activity
- ANSI appears ready to support working group
- Are You?