Mass Storage

Hui Chen [a]

[a]CUNY Brooklyn College

May 4, 2020

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Physical Structures

The main mass-storage system in modern computers is secondary storage.

- Hard disk drives (HDD)
- I Nonvolatile memory (NVM) devices

In addition, there are,

- RAM drives
- Magnetic tapes

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Performance Characteristics of HDD

- Disk rotation speed
- Transfer rate and effective transfer rate
- Positioning time (also called random-access time)
  - Seek time
  - Rotational latency
- Head crash

Performance Characteristics of NVM

- No mechanical moving parts, no seek time or rotational latency, and less power consumption
- NVM can be “worn”
  - Life span and drive writes per day (DWPD)
  - Over-provisioning space and wear leveling
- Cannot be overwritten, can only be eraesed.
  - Flash translation layer (FTL)
  - Page and block (read/write pages, erase blocks)
  - Gargabe collection
  - Write Amplification

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Address Mapping

- Storage devices are addressed as large one-dimensional arrays of logical blocks
  - The logical block is the smallest unit of transfer
  - The address of a block is a Logical Block Address (LBA)
- The one-dimensional array of logical blocks is mapped onto the sectors or pages of the device, e.g.,
  - HDD. CHS (cylinder, head, sector) → LBA
  - NVM. CBP (chip, block, page) → LBA

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Connection Methods

- I/O bus, e.g.,
  - Advanced Technology Attachment (ATA), Serial ATA (SATA), eSATA, Serial Attached SCSI (SAS), Universal Serial Bus (USB), and Fibre Dhannel (FC).
- Host Bus Adapter (HBA)
  - Host controller
  - Device controller

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

HDD Scheduling

Also called disk (or disk arm, or disk head) scheduling, primarily on minimizing the amount of disk head movement.

- First-come, first-served scheduling (FCFS)
- SCAN scheduling (also the elevator algorithm)
- C-SCAN scheduling

NVM Scheduling

SSD schedulers have exploited a few properties, such as,

- write service time is not uniform, and
- NVM can be “worn”.

HDD and NVM Scheduling on Linux

On Linux, there are

- the Deadline scheduling
- the NOOP scheduling
- the Completely Fair Queueing scheduling (CFQ)

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Partition, Volumes, and Formatting

1. Low-level formatting (i.e., physical formatting)
   - Error detection and correction (e.g., checksums, parity; hamming code)
2. Partition creation
   - Boot block (boot sector), bootable, non-bootable.
3. Volume creation
4. Logical formatting and file system creation
   - Blocks vs. clusters?
5. Mounting file system and volumes
6. Raw disks
7. Dealing with bad blocks (sector sparing or forwarding; sector slipping for HDDs)

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Managing Swap-Space

For memory management, needs seconardy storage for swapping and paging

- size?
- location?
- how?

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Storage Attachment

- Host-attached storage
- Network-attached storage (NAS)
- Cloud storage
- Storage-area networks (SAN) and storage arrays

Outline

1. Physical Structures
2. Performance Characteristics
3. Addressing
4. Connect to Host
5. Scheduling for Secondary Storage
6. Storage Device Management
7. Swap-Space Management
8. Storage Attachment
9. Storage Reliability and Performance

Redundancy and Performance

- Reliability via storage redundancy.
- Peformnance via I/O parallelsim.

RAID

- Level 0. Striping only, no mirroring or parity.
- Level 1. Mirroring, without parity or striping.
- Level 1+0 and 0+1.
- Level 2. (not used) Bit-level striping with dedicated Hamming-code parity.
- Level 3. (rarely used) Byte-level striping with dedicated parity.
- Level 4. Block-level striping with dedicated parity.
- Level 5. Block-level striping with distributed parity.
- Level 6. P+Q redundancy (block-level striping with double distributed parity.)

Extension to RAID, e.g., ZFS

- Blocks are checksumed to reduce software errors.
- Combine both volume and file system management (pool of storage).