Mass-Storage Systems

Unit-IV

Lecture -1

March 12, 2020

Session Objectives

- To describe the physical structure of secondary storage devices and its effects on the uses of the devices
- To explain the performance characteristics of mass-storage devices
- To evaluate disk scheduling algorithms
- To discuss operating-system services provided for mass storage, including RAID

Session Outcomes

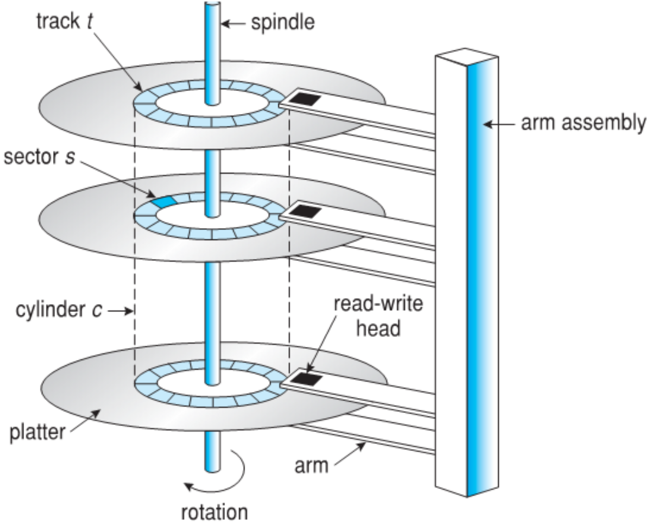
At the end of this session, participants will be able to

- Discuss the Mass Storage Structure
  Disk Structure
  Disk Attachment
  Disk Scheduling
  Disk Management

Overview of Mass Storage Structure

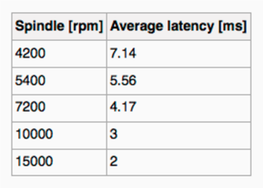
- **Magnetic disks** provide bulk of secondary storage of modern computers
  - Drives rotate at 60 to 250 times per second
  - **Transfer rate** is rate at which data flow between drive and computer
  - **Positioning time** (**random-access time**) is time to move disk arm to desired cylinder (**seek time**) and time for desired sector to rotate under the disk head (**rotational latency**)
  - **Head crash** results from disk head making contact with the disk surface -- That’s bad
- Disks can be removable
- Drive attached to computer via **I/O bus**
  - **Host controller** in computer uses bus to talk to **disk controller** built into drive or storage array

Moving-head Disk Mechanism



Hard Disks

- Platters range from .85” to 14” (historically)
  - Commonly 3.5”, 2.5”, and 1.8”
- Range from 30GB to 3TB per drive
- Performance
  - Transfer Rate – theoretical – 6 Gb/sec
  - Effective Transfer Rate – real – 1Gb/sec
  - Seek time from 3ms to 12ms – 9ms common for desktop drives
  - Average seek time measured or calculated based on 1/3 of tracks
  - Latency based on spindle speed
    - 1 / (RPM / 60) = 60 / RPM
  - Average latency = ½ latency



Spindle [rpm]	Average latency [ms]
4200	7.14
5400	5.56
7200	4.17
10000	3
15000	2

(From Wikipedia)

The First Commercial Disk Drive

1956
IBM RAMDAC computer
included the IBM Model
350 disk storage system

5M (7 bit) characters
50 x 24” platters
Access time = < 1 second

Solid-State Disks

- Non volatile memory used like a hard drive
- Many technology variations
- Can be more reliable than HDDs
- More expensive per MB
- Maybe have shorter life span
- Less capacity
- But much faster
- No moving parts, so no seek time or rotational latency

Magnetic Tape

- Was early secondary-storage medium
- Relatively permanent and holds large quantities of data
- **Access time slow**
- Random access 1000 times slower than disk
- Mainly used **for backup, storage of infrequently-used data**, transfer medium between systems
- Once data under head, transfer rates comparable to disk
- 140MB/sec and greater
- 200GB to 1.5TB typical storage

Disk Structure

- Disk drives are addressed as large 1-dimensional arrays of **logical blocks**, where the logical block is the smallest unit of transfer
  - Low-level formatting creates **logical blocks** on physical media
- The 1-dimensional array of logical blocks is mapped into the sectors of the disk sequentially
  - Sector 0 is the first sector of the first track on the outermost cylinder
  - Mapping proceeds in order through that track, then the rest of the tracks in that cylinder, and then through the rest of the cylinders from outermost to innermost

Disk Attachment

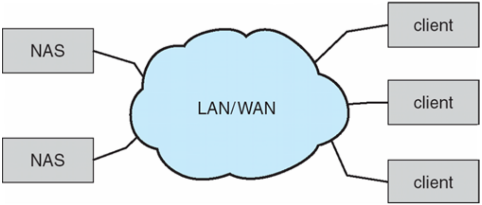
- Computers access disk storage in two ways.
- One way is via I/O ports (or **host-attached storage**); this is common on small systems.
- The other way is via a remote host in a distributed file system;
  this is referred to as **network-attached storage**.

Host-Attached Storage

- Host-attached storage accessed through I/O ports talking to I/O busses
- SCSI itself is a bus, up to 16 devices on one cable, **SCSI initiator** requests operation and **SCSI targets** perform tasks
- FC is high-speed serial architecture
- Have 24-bit address space – the basis of storage area networks (SANs) in which many hosts attach to many storage units
- I/O directed to bus ID, device ID, logical unit (LUN)

Network-Attached Storage

- Network-attached storage (NAS) is storage made available over a network rather than over a local connection (such as a bus)
- Remotely attaching to file systems
- NFS and CIFS are common protocols
- Implemented via remote procedure calls (RPCs) between host and storage over typically TCP or UDP on IP network
- iSCSI protocol uses IP network to carry the SCSI protocol
- Remotely attaching to devices (blocks)



Network-Attached Storage

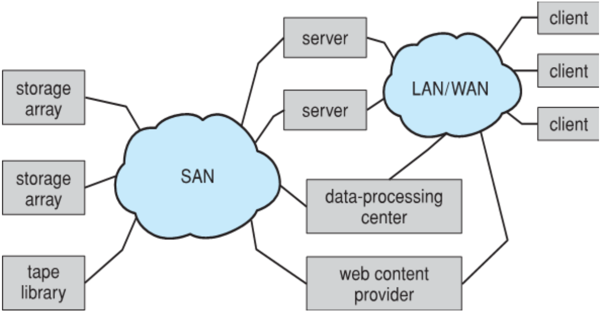
- One drawback of network-attached storage systems is : that the storage I/O operations **consume bandwidth** on the data network,
- Thereby **increasing the latency** of network communication.
- This problem is more noticeable more in **large client–server** installations
  the communication between servers and clients **competes for bandwidth with the communication** among servers and storage devices.

Storage Array

- Can just attach disks, or arrays of disks
- Storage Array has controller(s), provides features to attached host(s)
- Ports to connect hosts to array
- A few to thousands of disks
- RAID, hot spares, hot swap
- Shared storage $->$ more efficiency

Storage Area Network

- A storage-area network (SAN) is a private network (using storage protocols rather than networking protocols) connecting servers and storage units
- Common in large storage environments
- Multiple hosts attached to multiple storage arrays - flexible



Storage Area Network

- SAN is one or more storage arrays
- Connected to one or more Fibre Channel switches
- Hosts also attach to the switches
- Easy to add or remove storage, add new host and allocate it storage
- Over low-latency Fibre Channel fabric

Disk Scheduling

- The operating system is responsible for using hardware efficiently — for the disk drives, this means having a **fast access time and disk bandwidth ;Minimize seek time**
- The seek time is the time for the disk arm to move the heads to the cylinder containing the desired sector.
- The rotational latency is the additional time for the disk to rotate the desired sector to the disk head. The
- Disk bandwidth is the total number of **bytes transferred**, divided by the **total time between the first request for service and the completion of the last transfer**

Disk Scheduling

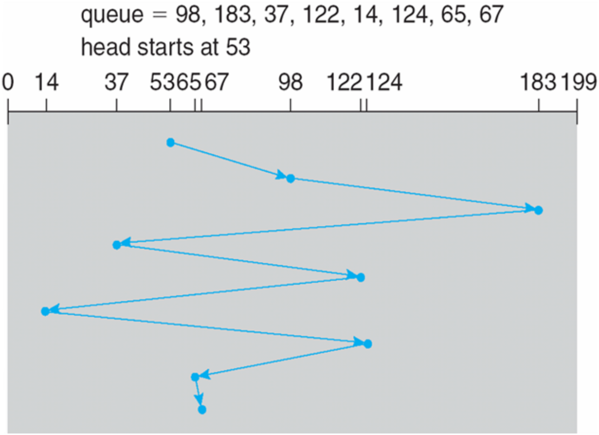
- There are many sources of disk I/O request OS
- System processes ; Users processes
- I/O request includes input or output mode, disk address, memory address, number of sectors to transfer
- OS maintains queue of requests, per disk or device
- Idle disk can immediately work on I/O request, busy disk means work must queue
- Optimization algorithms only make sense when a queue exists

Disk Scheduling

- Note that drive controllers have small buffers and can manage a queue of I/O requests
- Several algorithms exist to schedule the servicing of disk I/O requests
- We illustrate scheduling algorithms with a request queue (0-199) 98, 183, 37, 122, 14, 124, 65, 67
- Head pointer 53

FCFS

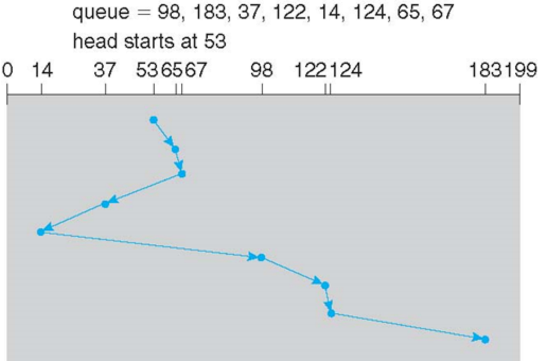
Illustration shows total head movement of 640 cylinders



does not provide the fastest service

SSTF

- Shortest Seek Time First selects the request with the minimum seek time from the current head position
- This is a form of SJF scheduling; may cause starvation
- Illustration shows total head movement of 236 cylinders

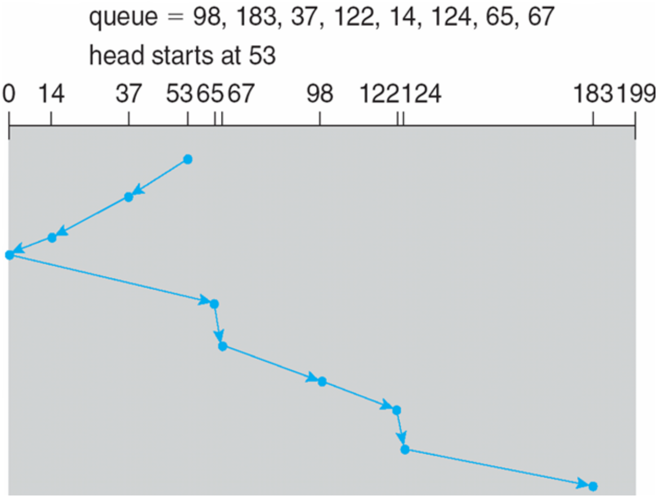


it is not optimal

SCAN

- The disk arm starts at one end of the disk, and moves toward the other end, servicing requests until it gets to the other end of the disk, where the head movement is reversed and servicing continues.
- SCAN algorithm Sometimes called the elevator algorithm
- Illustration shows total head movement of 236 cylinders
- But note that if requests are uniformly dense, largest density at other end of disk and those wait the longest

SCAN



queue = 98, 183, 37, 122, 14, 124, 65, 67

head starts at 53

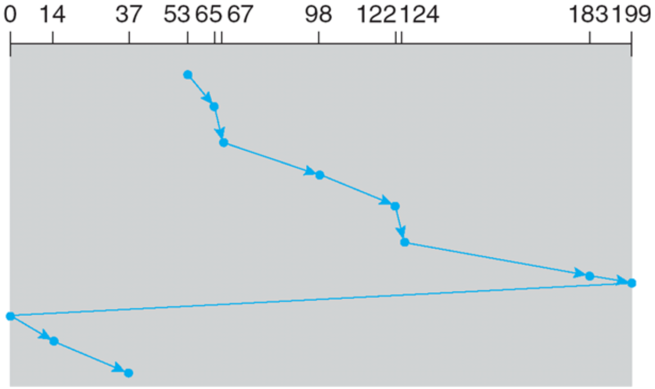
C-SCAN

- Provides a more uniform wait time than SCAN
- The head moves from one end of the disk to the other, servicing requests as it goes
- When it reaches the other end, however, it immediately returns to the beginning of the disk, without servicing any requests on the return trip
- Treats the cylinders as a circular list that wraps around from the last cylinder to the first one

C-SCAN

queue = 98, 183, 37, 122, 14, 124, 65, 67

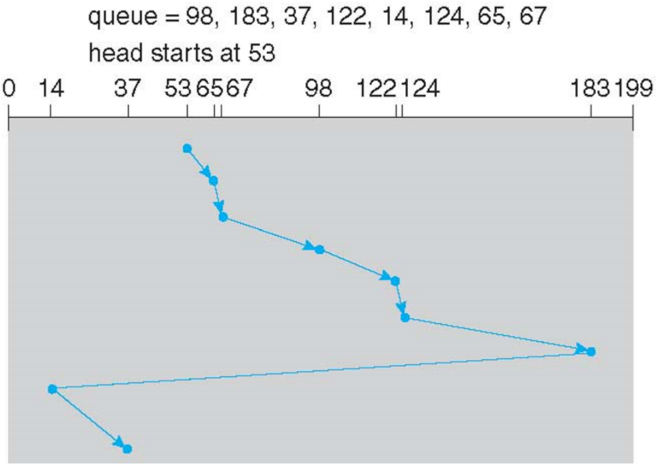
head starts at 53



C-LOOK

- LOOK a version of SCAN, C-LOOK a version of C-SCAN
- Arm only goes as far as the last request in each direction, then reverses direction immediately, without first going all the way to the end of the disk

C-LOOK



queue = 98, 183, 37, 122, 14, 124, 65, 67

head starts at 53

Selecting a Disk-Scheduling Algorithm

- **SSTF is common** and has a natural appeal
- **SCAN and C-SCAN** perform better for systems that place a **heavy load** on the disk
- Less starvation
- Performance depends on the number and types of requests
- The disk-scheduling algorithm should be written as a separate module of the operating system, allowing it to be replaced with a different algorithm if necessary
- Either SSTF or LOOK is a reasonable choice for the default algorithm

Disk Management

- **Low-level formatting, or physical formatting —** Dividing a disk into sectors that the disk controller can read and write
- Each sector can hold header information, plus data, plus **error correction code (ECC)**
- Usually 512 bytes of data but can be selectable
- To use a disk to hold files, the operating system still needs to record its own data structures on the disk
- Partition the disk into one or more groups of cylinders, each treated as a logical disk
- **Logical formatting** or "making a file system"

Disk Management

- Boot block initializes system
- The bootstrap is stored in ROM
- **Bootstrap loader** program stored in boot blocks of boot partition
- Methods such as **sector sparing** used to handle bad blocks