SECONDARY STORAGE DEVICES: MAGNETIC TAPES AND CD-ROM

Contents of today's lecture:

- Magnetic Tapes
  - Characteristics of magnetic tapes
  - Data organization on 9-track tapes
  - Estimating tape length requirements
  - Estimating data transmission times
  - Disk versus tape
- CD-ROM
  - Physical Organization of CD-ROM
  - CD-ROM Strengths and Weaknesses

Reference: Folk, Zoellick and Riccardi, File Structures, 1998. Section 3.2, 3.5 and 3.6.

Characteristics of Magnetic Tapes

- No direct access, but very fast sequential access.
- Resistant to different environmental conditions.
- Easy to transport, store, cheaper than disk.
- Before, it was widely used to store application data; nowadays, it's mostly used for backups or archives (tertiary storage).

Data Organization on Nine-Track Tapes

In a tape, the **logical position** of a byte within a file is the same as its **physical position** in the file (sequential access).

Nine-track tape:

```
-----------------------------------------------------
______|0     0|__________|0|__________|0     0|______
______|0     0|__________|1|__________|0     0|______
______|0     0|__________|1|__________|0     0|______
______|0     0|__________|0|__________|0     0|______
______|0     0|__________|1|__________|0     0|______
______|0     0|__________|0|__________|0     0|______
______|0     0|__________|0|__________|0     0|______
______|0     0|__________|1|__________|0     0|______
______|0     0|__________|1|__________|0     0|______
-----------------------------------------------------
      |<-Gap->|  <-- Data Block -->   |<-Gap->|
```

- **Data blocks** are separated by interblock GAPS.
- 9 parallel tracks (each is a sequence of bits)
- A **frame** is a 1-bit slice of the tape corresponding to 9 bits (one in each track) which correspond to 1 byte plus a **parity bit**.

In the example above, the byte stored in the frame that is shown is: `01101001`. The parity bit is `1`, since we are using **odd parity**, i.e., the total number of bits is odd.

Complete the parity bit in the examples below:

11111111

00000000

00100000

Since 000000000 cannot correspond to a valid byte, this is used to mark the **interblock gap**.

So, if we say that this tape has 6,250 **bits per inch** (bpi) per track, indeed it stores 6,250 **bytes per inch** when we take into account the 9 tracks.

Estimating Tape Length Requirements

Performance of tape drives can be measured in terms of 3 quantities:
- Tape density = 6250 bpi (bits per inch per track)
- Tape speed = 200 inches per second (ips)
- Size of interblock gap = 0.3 inch

File characteristics:
- Number of records = 1,000,000
- Size of record = 100 bytes

How much tape is needed?

It depends on the blocking factor (how many records per data block). Let us compute the space requirement in two cases:
A) Blocking factor = 1
B) Blocking factor = 50

Space requirement (s)
b = length of data block (in inches)
g = length of interblock gap (in inches)
n = number of data blocks

s = n x (b + g)

A) Blocking factor = 1

b = block size/tape density = 100 bytes/6250 bpi= 0.016 inch
n = 1,000,000 (recall blocking factor is 1)
s = 1,000,000 x (0.016 + 0.3) inch = 316,000 inches $\sim$ 26,333 feet
(Absurd to have the length of the data block smaller than the interblock gap!)

B) Blocking factor = 50

b = 50 x 100 bytes/6,250 bpi = 0.8 inch
n = 1,000,000 records/50 records per block = 20,000 blocks
s = 20,000 x (0.8 + 0.3) inch = 22,000 inches $\approx$ 1,833 feet

An enormous saving by just choosing a higher blocking factor.

Effective Recording Density (ERD)

ERD = number of bytes per block / number of inches to store a block

In previous example :

A) Blocking factor =1: E.R.D. = 100/0.316 $\sim$ 316.4 bpi
B) Blocking factor =50: E.R.D. = 5,000/1.1 $\sim$ 4,545.4 bpi

The **Nominal Density** was 6,250 bpi!

Estimating Data Transmission Times

Nominal Rate = tape density (bpi) x tape speed (ips)

In a 6,250 - bpi , 200 - ips tape :
Nominal Rate = 6,250 bytes/inch x 200 inches/second =
= 1,250,000 bytes/sec $\sim$ 1,250 KB/sec

Effective Transmission Rate = E.R.D. x tape speed

In the previous example:
A) E.T.R. = 316.4 bytes/inch x 200 inches/sec = 63,280 bytes/sec $\sim$ 63.3 KB/sec
B) E.T.R. = 4,545.4 bytes/inch x 200 inches/sec = 909,080 bytes/sec $\sim$ 909 KB/sec
Note : There is a tradeoff between **increasing** blocking factor for increasing speed & space utilization and **decreasing** it for reducing the size of the I/O buffer.

Disk versus Tape

In the past : Disks and Tapes were used for secondary storage: disks preferred for random access and tapes for sequential access.
Now :
Disks have taken over most of secondary storage (lower cost of disk and lower cost of RAM which allows large I/O buffer). Tapes are mostly used for **tertiary storage**.

Physical Organization of CD-ROM

Compact Disc - read only memory (write once)

- Data is encoded and read optically with a laser
- Can store around 600 MB data

Digital data is represented as a series of **Pits** and **Lands**.

Pit = a little depression, forming a lower level in the track
Land = the flat part between pits, or the upper levels in the track

Reading a CD is done by shining a laser at the disc and detecting changing reflections patterns.

1 = change in height (land to pit or pit to land)
0 = a "fixed" amount of time between 1's

```
       LAND       PIT         LAND       PIT    LAND
... ------------+     +-------------+       +---- ...
                |_____|             |_______|

... 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 ...
```

Changes in height in the track are detected as changes of intensity of the reflected light.
Note: We cannot have two 1's in a row!

Indeed, because of other limitations there must be at least two and at most ten 0's between two 1's.

Therefore, each of the 256 bytes must be encoded into a sequence of bits that has every pair of 1's separated by at least two zeros. There are exactly 267 binary words of length 14 that satisfy this property; 256 of them were chosen to represent every possible byte in the so-called eight to fourteen modulation. We could not encode bytes using 13 bits since there are only 188 words of length 13 having the desired property.

Eight to fourteen modulation (EFM) encoding table:

Decimal Value	Original Bits	Translated Bits
0	00000000	01001000100000
1	00000001	10000100000000
2	00000010	10010000100000
3	00000011	10001000100000
4	00000100	01000100000000
5	00000101	00000100010000
6	00000110	00010000100000
7	00000111	00100100000000
8	00001000	01001001000000
...	...	...

Note that: Since 0's are represented by the **length of time** between transitions, we must travel at **constant linear velocity** on the tracks.

Comparing CD-ROM with magnetic disks

CR-ROM	**Magnetic Disks**
CLV = Constant Linear Velocity	CAV = Constant Angular Velocity
Sectors organized along a spiral	Sectors organized in concentric track
Sectors have same linear length (data packed at its maximum density permitted)	Sectors have same angular length (data written less densely in the outer tracks)
Advantage: takes advantage of all storage space available	Advantage: operates on constant speed, timing marks to delimit tracks
Disadvantage: has to change rotational speed when seeking (slower towards the outside)	Disadvantage: it doesn't use up all storage available

CD-ROM addressing and poor Seek performance

Addressing

1 second of play time is divided up into 75 **sectors**.
Each sector holds 2KB.

60 Min CD :
60 min x 60 sec/min x 75 sectors/sec = 270,000 sectors = 540,000 KB $\sim$ 540 MB

A **sector** is addressed by :
Minute : Second : Sector
16:22:34
16 min, 22 sec, 34th sector

Difficulty in Seeking

- To read address of a sector it must be at the correct speed
- But knowing the correct speed depends on the ability to read the address info!

The **drive control mechanism** solves this problem by trial-and-error.
This slows down the performance!

CD-ROM Strength and Weaknesses

- Seek performance ($\sim$ 500 msecs) - Slow

  Our old analogy :
  20 secs (RAM)
  58 days (Magnetic Disks)
  2.5 years (CD-ROM)

- Data transfer rate - 150 KB/sec - Slow (while $\sim$ 3,000 KB/sec for magnetic disks), but 5 times faster than floppy disks.

- Storage capacity is $\sim$ 600 MB; good for storing texts.

- Read-only access (publishing medium). File structure designer can take advantage of that.

Things changed nowadays :

- Most drives use CAV or combination of CAV and CLV
- Other types of compact discs :
  - CD-R = compact disc-recordable
  - CD-RW = compact disc-rewritable
  They use different technologies which simulates the effect of Pits and Lands.