CS245 Database System Principles - Winter 2001

Assignment 1

- Due in class on Thursday, Jan 18.
- State all assumptions.
- Subscribe to **`cs245-win01`** to receive clarifications and changes.
- Email questions to **`cs245ta-win01@lists.stanford.edu`**

Problem 1 (40 points)

Consider a 3.5 inch diskette with 2 magnetic surfaces with 32 tracks each, rotating at 3600 rpm. It has a usable capacity of 1.5 megabytes (1.5×2^{20} bytes). Assume 20% of each track is used as overhead.

a. What is the burst bandwidth this disk could support reading a single block from one track?
b. What is the sustained bandwidth this disk could support reading an entire track?
c. What is the average rotational latency, assuming it is not necessary to start at the beginning of the track?
d. Assuming the average seek time is 18 ms, what is the average time to fetch a 2-kilobyte (2^{11} bytes) sector?

Problem 2 (15 points)

Consider the Megatron 747 disk with the following properties:

- There are four platters providing eight surfaces.
- There are 2^{13}, or 8192 tracks per surface.
- There are (on average) $2^8 = 256$ sectors per track.
- There are $2^9 = 512$ bytes per sector.
- The disk rotates at 3840 rpm.
- The block size is $2^{12} = 4096$ bytes.
- Assume 10% of each track is used as overhead.
- The time it takes the head to move *n* tracks is *1 + n/500* milliseconds.

Suppose that we know that the last I/O request accessed cylinder 3000.

a. What is the expected (average) number of cylinders that will be traveled due the very next I/O request to this disk?
b. What is the expected block access time for the next I/O, again given that the head is on cylinder 3000 initially?

Problem 3 (15 points)

Suppose that we are scheduling I/O requests for the new Megatron 747 disk (Problem 2). Recall that the average seek, latency and transfer times are 6.5, 7.8, and 0.5 milliseconds. Initially the head is at cylinder 4000, and then the following requests come in:

- time = 0 milliseconds; request for block on cylinder 7000 arrives
- time = 3 milliseconds; request for block on cylinder 2000 arrives
- time = 11 milliseconds; request for block on cylinder 7500 arrives
- time = 19 milliseconds; request for block on cylinder 3000 arrives

a. If we use the elevator scheduling algorithm, at what time is each request serviced completely?
b. If we use a first-come-first-served scheduler, at what time is each request serviced fully?

Problem 4 (10 points)

What are the advantages and disadvantages of using fixed-length records of fixed-length fields. Give at least 2 points for each and justify each point you give.

Problem 5 (20 points)

You are designing a file system for a medical application. Each patient record has 12 fields that always occur (e.g., name, patient number) and 38 fields that may or may not be relevant or known for a patient (e.g., number of children given birth to, cholesterol level). Assume that each of the optional fields is relevant or known for a particular patient with probability *p*. For all fields, values are stored in a fixed 15 bytes.

You are considering two options:
(i) A fixed format record.
(ii) A variable format record where all optional fields are tagged. Each tag is two bytes.

a. What is the expected size of a record for each option? (Your answer may be a function of *p*.)
b. For what range of p values is the fixed format option best?
c. Given that $p=0.8$, which option is better?