CS-245 Database System Principles – Winter 2002
Assignment 1

Due in class on Tuesday, January 22.

- State all assumptions.
- Subscribe to **cs245@lists.stanford.edu** to receive clarifications and changes.
- You can email questions to **cs245-staff@cs.stanford.edu**.

Problem 1 (40 points)

Consider a 3.5 inch diskette with 2 magnetic surfaces. Each surface has 64 tracks each. The disk rotates at 5400 RPM, and 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 15 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 $2^{13} = 8192$, 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 2000.

a. What is the expected (average) number of cylinders that will be traveled due to 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 2000 initially?

Problem 3 (15 points)

Suppose that we are scheduling I/O requests for the new Megatron 747 disk (described in problem 2). Recall that the average latency and transfer times are 7.8 and 0.5 milliseconds (respectively), and that the seek time is (1+n/500) milliseconds. Initially the head is at cylinder 5000, and then the following requests come in:

	Time	Request
1.	0 milliseconds	Request for block on cylinder 7000
2.	3 milliseconds	Request for block on cylinder 2000
3.	11 milliseconds	Request for block on cylinder 7500
4.	19 milliseconds	Request for block on cylinder 3000
5.	26 milliseconds	Request for block on cylinder 4500

a. If we use the elevator scheduling algorithm, state for each request what time it is serviced completely.
b. If we use a first-come-first-served scheduler, state for each request what time it is serviced completely.

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 filesystem for a medical application. Each patient record has 10 fields that always occur (e.g., name, patient number) and 34 fields that may or may not be relevant or known for a patient (e.g., number of children given birth to, cholesterol level, etc). Assume that each of the optional fields is relevant or known for a particular patient with probability p. For the required fields, the values are stored in a fixed 30 bytes. For the optional fields, the values are stored in a fixed 10 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. The sequence (possibly empty) of optional fields is terminated with a special two byte tag.

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?