

Bilkent University
Department of Computer Engineering
CS351 DATA ORGANIZATION AND MANAGEMENT

Midterm Exam Date: November 7, 2009 Time: 13:15 – 15:15

Student Name/ ID No. Section	SOLUTIONS
---------------------------------	-----------

GOOD LUCK!

- Notes:
1. There are 100 points, 6 questions on 6 pages.
 2. Please READ the questions. It is a closed book/notes exam.
 3. Show your work.
 4. You are not allowed to use your cell phone or PDA for any purpose.
 5. You cannot leave the exam room in the first 30 minutes.
 6. Please do not write anything inside the following table.

Q. No.	Q1	Q2	Q3	Q4	Q5	Q6	Overall
Points Possible	18	12	12	18	20	20	100
Your Grade							

1. (18%) Consider a file of 120 Mbytes with record size 100 bytes stored on a disk drive with the disk parameters given below:
 $s=20$ msec, $r=10$ msec, $btt=1.00$ msec, $ebt=1.10$ msec. and block size 2400 bytes.
 - a) Compute T_x (Exhaustive reading of the entire file.)
 - b) Compute T_F (Average time to Fetch or Access a single record)
 - c) How many records can be read randomly in time T_x ?

$$B = \frac{120000000}{100 \times 2400} = 50000$$

$$a) T_x = b \cdot ebt = 50000 \times 1.1 = 55000 \text{ msec} = 55 \text{ sec}$$

$$b) T_F = \frac{T_x}{2} = 27.500 \text{ msec} = 27.5 \text{ sec}$$

c) 2

2. (12%) Consider a sequential magnetic medium (tape) with the following parameters:

1600 bytes/inch recording density

0.5 inches Inter Block Gap

2500 feet tape length of a reel (30000 inches)

10 msec. of start/stop time

200 inches/sec read/write speed

a) Does a file with 60 000 records of 400 bytes each fit on a single reel if the blocking is 1?

b) What is the minimum Blocking Factor for which this file fits on a single reel of tape?

a)

$$60000 \times 400 = 24,000,000 \text{ byte}$$

$$\text{Record length} = \frac{24,000,000}{1600} = 15,000 \text{ inch}$$

$$\text{IBG} = 60000 \times 0.5 = 30,000$$

$$\text{Total} = 45,000 \text{ inch} > 30,000 \text{ inch} \Rightarrow \text{NO}$$

$$b) \quad 15000 + \frac{60000 \times 0.5}{B} = 30000 \quad B = 2$$

4. (18%) Consider a file of 120 Mbytes with record size 100 bytes stored on a disk drive with the disk parameters given below:

$s=20$ msec, $r=10$ msec, $btt=1.00$ msec, $ebt=1.10$ msec. and block size 2400 bytes.

- In the first phase of a Sort-Merge, the 120 Mbyte file is sorted using Replacement Selection with a 2 Mbyte RAM and two independent disk drives to produce 40 initial strings approximately 3 Mbytes each. What is the best time in which this phase can be completed?
- If two-way merges are used in the merge phase, how many passes are needed to merge the 40 strings?
- Compute the time to merge two initial strings (each of size 3 Mbytes) using 2 Mbytes of RAM and a single disk drive.

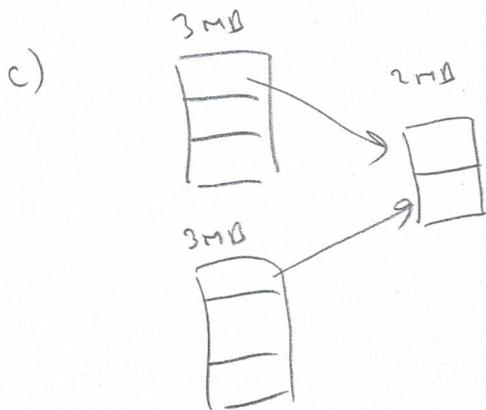
$$a) \quad b = \frac{120 \times 10^6}{2400} = 50,000$$

Using replacement selection sort we can overlap I/O

\therefore only find the input time.

$$b \times ebt = 50,000 \times 1.10 = 55,000 \text{ msec} = 55 \text{ sec.}$$

$$b) \quad \lceil \log_2 40 \rceil = 6$$



$R, R, W, R, W, R, W, R, W, R, W$

$$6R + 5W \Rightarrow 11 \text{ I/O activity}$$

$$11 \times (s+r) =$$

$$11 \times (20+10) = 330 \text{ msec}$$

$$\text{No. of block} = \frac{6 \times 10^6}{2400} = 2500$$

$$2 \times 2500 \times 1.10 = 5500 \text{ msec}$$

↑
R & W

$$\begin{array}{r} 5500 \\ + 330 \\ \hline 5830 \text{ msec} \Rightarrow 5.8 \text{ sec} \end{array}$$

5. (20%) In Linear Hashing, the boundary value b_v indicates the address of the disk bucket which will be split next, following the net addition of $Bkfr * Lf$ records to the file in order to maintain a constant load factor (Lf). $Bkfr$ denotes the bucket factor, the number of records that fit in a disk bucket.

If $b_v = 41$, specify:

- The minimum number of bits, k , (or minimum Hashing Level h) needed to address the buckets that are already split.
- The number of buckets of the primary area that are yet to split until the next Hashing Level is reached.
- The total number of buckets in the primary area of the file (excluding the overflow buckets, if any).
- If the Load Factor is $2/3$ and the bucket capacity is 6 records:
 - What is the possible minimum number of records in the file for $b_v = 41$?
 - What is the possible maximum number of records in the file for $b_v = 41$?

a) $b_v = 41 = \underbrace{101001}_{6 \text{ digit}} \quad h = 6$
 $\Rightarrow \underline{7} \text{ digit} = \text{min \# of}$

b)

$h = 6$

$b_v \rightarrow \left[\begin{array}{l} 41 \\ \vdots \\ 63 \end{array} \right]$

$2^6 - 1$

$63 - 41 + 1 = \underline{23}$

c)

$\left. \begin{array}{l} 0 \\ 40 \\ 41 \end{array} \right\} \begin{array}{l} h+1 \\ 41 \text{ base} \end{array}$

$\left. \begin{array}{l} 63 \\ 64 \end{array} \right\} \begin{array}{l} 23 \text{ base} \\ 41 \text{ base} \end{array}$

104

$41 + 41 + 23 = \underline{105}$

d) 1) $105 \times 6 \times \frac{2}{3} = \underline{420}$

$104 \times 6 \times \frac{2}{3} + 4 = 416 + 4 = \underline{420}$

2) $6 \times \frac{2}{3} = 4$ 3 elements $\Rightarrow \underline{423}$

37 = 5 = 00101
 22 = 22 = 10110
 51 = 19 = 10011

42 = 10 = 01010
 21 = 21 = 10101
 17 = 17 = 10001
 12 = 12 = 01100

35 = 3 = 00011
 36 = 4 = 00100
 56 = 24 = 11000

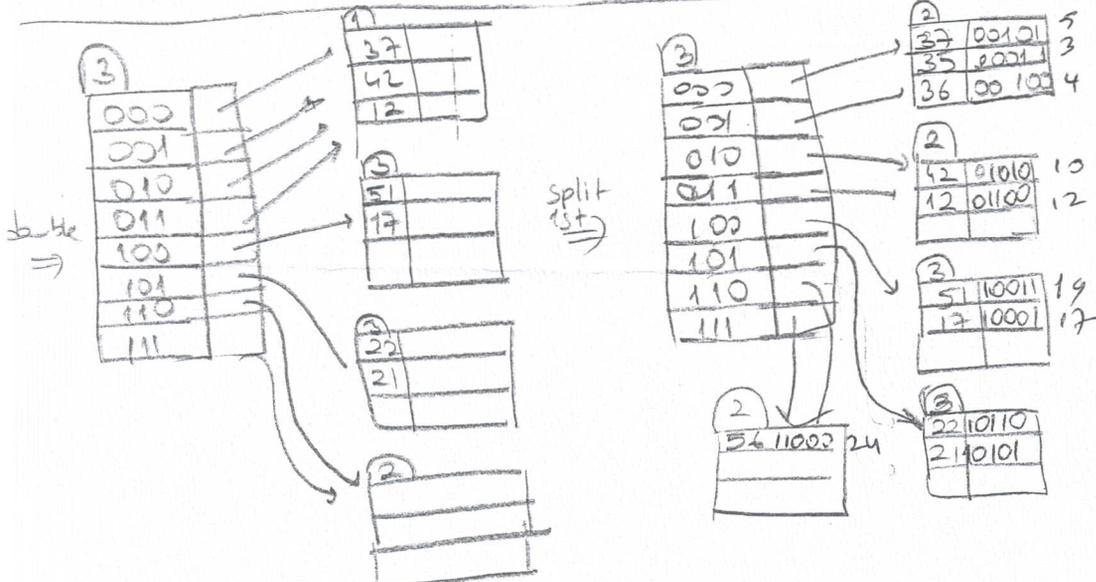
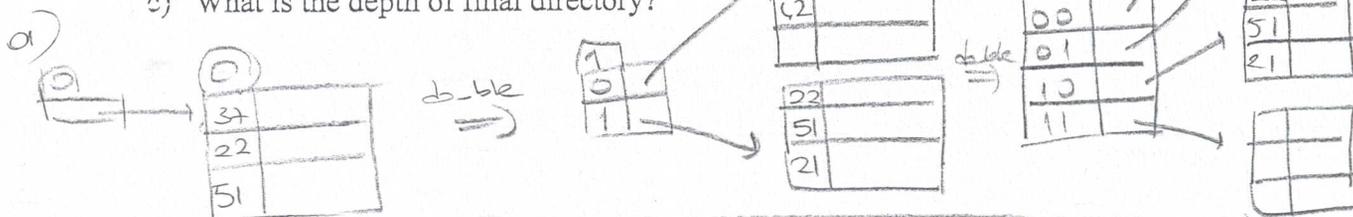
6. (20%) Consider the following key values:

37 22 51 42 21 17 12 / 35 36 56

a) Using a hash function $H(x) = x \text{ MOD } 32$ and 5-bit pseudokeys, create the directory and the disk pages for Extensible Hashing by successively inserting the keys in the order given. Use the leftmost bits from the pseudokeys in addressing. Assume page capacity is 3 records. Show all the page splits and the doubling of the directory. Label the disk pages as A, B, C, in that order. Start with a depth 0 directory (with a single entry).

b) What is the load factor in the final configuration?

c) What is the depth of final directory?



b) $\frac{10}{5 \cdot 3} = \frac{10}{15} = \frac{2}{3}$

c) 3