SECOND MIDTERM

Please answer all 5 questions. Read carefully. We suggest reserving part of the exam book for your scratch work — work on these before attempting your actual answers. This is a closed book exam, but you may refer to TWO prepared 8"x11" 2-sided sheet of notes. Good Luck!

Problem 1 TRUE or FALSE [5 points each]

Minus 3 points for wrong answers. Brief justification is needed for full credit.

(a) Any comparison-based algorithm for computing the median must make at least $\lg n$ comparisons.

ANSWER: TRUE, by info theoretic argument, there must be at least n leaves in the comparison tree. This tree has height at least $\lg n$.

REMARKS: Of course, $\lg n$ is very far from the truth, since we can use other arguments to prove a lower bound of $n-1$. Roughly, any computation that do not look at every input number cannot guarantee to find the answer. This actually gives you a lower bound$n/2$ comparisons. How can you improve this to $n-1$?

(b) "*In a certain TRUE/FALSE question, you get 5 points for correct answer, 0 points for not attempting the question and -3 points for an incorrect guess. Suppose you have NO clue what the answer might be. It is a good idea to toss a coin and answer TRUE if head, and answer FALSE if tail.*"

ANSWER: TRUE. Your expected value is $(-3)(1/2) + (5)(1/2)$ which is equal to 1. This is better than the 0 points you would get if you do not attempt to answer the question.

Problem 2 RECURRENCES [15+20 points]

(a) Solve the recurrence

$$T(n) = 2T(n/2) + n \lg \lg n.$$

ANSWER: Use the transformation: $s(N) = T(2^N)/2^N$. Then $s(N) = s(N-1) + \lg N$. Hence $s(N) = \Theta(N \lg N)$ and $T(n) = n \cdot s(\ln n)$. Thus $T(n) = \Theta(n \lg n \lg \lg n)$.

REMARK: of course, our definition of $s(N)$ has simply combined the range transformation and domain transformation into one step.

(b) Prove the best upper bound you can for the following recurrence:

$$T(n) = n^2 + 8T(n/3) + 40T(n/6).$$

Use real induction. HINT: You may want to begin with *some* upper bound first, and see if you can improve. If you arrive at the equation $g(\alpha) = (8/3^\alpha) + (40/6^\alpha) = 1$, this is easily solved without calculators. Does case (0) of the Master Theorem suggest anything?

ANSWER: For a preliminary investigation of this recurrence, if you look at the term "$40T(n/6)$" alone, Master theorem would tell you that $T(n) = \Omega(n^{\log_6(40)})$. Note that $\log_6(40)$ is between 2 and 3.

Let us guess that $T(n) \le Cn^\alpha$ for some α that we will determine. Then, by applying real induction, we have

$$T(n) \le n^2 + 8 \cdot C \cdot (n/3)^\alpha + 40 \cdot C \cdot (n/6)^\alpha = n^2 + C \cdot n^\alpha \cdot g(\alpha) \tag{1}$$

where

$$g(\alpha) = \frac{8}{3^\alpha} + \frac{40}{6^\alpha}.$$

Note that $g(2) > 1$ and $g(3) < 1$. Suppose we choose α so that $g(\alpha) < 1$ and $\alpha > 2$. Then from equation (1), we have

$$T(n) \le Cn^\alpha \left(\frac{n^{2-\alpha}}{C} + g(\alpha) \right).$$

If we choose $C = 1/(1 - g(\alpha))$ then this implies

$$T(n) \leq Cn^{\alpha}$$

for all $n \geq 1$. This would verify our induction hypothesis.

REMARK: Let $\alpha = \alpha_0$ be the unique solution to the equation $g(\alpha) = 1$. The above argument shows that we can choose α to be $\alpha_0 + \epsilon$ for any $\epsilon > 0$. But, can you choose $\alpha = \alpha_0$? Well, case (0) of the Master theorem will suggest that you modify the induction hypothesis to

$$T(n) \leq Cn^{\alpha_0} \log n.$$

This can be verified also. You can even achieve

$$T(n) \leq Cn^{\alpha_0}$$

but you need to strengthen your induction hypothesis in another way (how?).

Problem 3 AVL TREES [15+20 points]

You need not show the keys, just draw the nodes in the following questions.
(a) How few nodes can there be in the smallest AVL tree of height 10?

ANSWER: We choose height $h = 10$ because it would be impossible to enumerate or draw the AVL tree. You need to work out the general principle. But we had already seen this principle in our lectures and notes. It is closely related to the Fibonacci sequence.

Here is the answer: Let $N(h)$ be the smallest number of nodes in an AVL tree of height h. Thus, $N(0) = 1$ and $N(1) = 2$. For $h \geq 1$, we have $N(h + 1) = 1 + N(h) + N(h - 1)$. Thus, $N(2) = 1 + 2 + 1 = 4$, which you can verify directly. Continuing,

$$N(3) = 7, N(4) = 12, N(5) = 20, N(6) = 33, N(7) = 54, N(8) = 88, N(9) = 143, N(232).$$

REMARK: you get no credit if you simply show me the number "232". How do I know that you have not gotten this number by magic, or worst? ALWAYS SHOW YOUR WORKING SOMEWHERE.
(b) Show an AVL tree of height 4, and a single deletion of a node in this tree that will result in 2 rotations.

ANSWER: Construct a tree of height 4 with $N(4) = 12$ nodes. Note that $N(4)$ is the minimum number in part (a). It is not necessary to use this minimum number, but it is the simplest example. Let u be the root of this tree, and u_L and u_R be its two children. Let the heights of u_L and u_R be 3 and 2, respectively. There is a leaf x below u_R of depth 2. Delete x. This will cause a double rotation below u_R. Then, you need one further rotation (either single or double, depending on the heights of the children of u_L.)

REMARKS: I intended the 2 rotations to be either a "single rotation" or a "double rotation". Since I did not specify that "double rotation" counts as "one rotation", I gave you the benefit of the doubt. If you only use a height 3 tree, which some of you did, you cannot achieve the desire effect.

Problem 4 (a, b)-TREES [10+20 points]

(a) What is the minimum number of nodes in an (a, b)-tree of height h?

ANSWER: Each non-leaf must have at least a children, except the root can have only 2 children. Hence the number of nodes at level $\ell = 0, 1, 2, 3, 4, \ldots$ (respectively) is at least

$$1, 2, 2a, 2a^2, 2a^3, \ldots.$$

Summing from levels 0 to level ℓ, we get

$$1+2\sum_{i=0}^{\ell-1} a^i = 1+2\frac{a^\ell-1}{a-1}.$$

REMARKS. You get partial credit if you do not know how to put the summation in closed form. The most frequent error is not remembering that the root may have only 2 children – but you can get most of the credit even in this case.
(b) In our homework, we discussed a method for splitting and merging ("fusion of") nodes in an (a,b)-tree to achieve an utilization ratio of a/b that is more than $2/3$. Generalize the arguments to achieve a ratio of $m/(m+1)$ for any integer m. You must state the needed inequalities on a and b.

ANSWER: You need to ensure that $a \geq m+1$. Then, when you need to merge or split, you can be assured of at least $a-1 \geq m$ siblings in which to borrow (if you are short) or share (if you are long) your keys.

In case of merging, your current node has $a-1$ keys. We may assume that the other m siblings has a keys each. We can combine all these $a(m+1)-1$ keys and split them into m new nodes. This requires

$$a \leq \frac{mb+1}{m+1}. \qquad (2)$$

In case of splitting, your current node has $b+1$ keys. You may assume that $m-1$ siblings with b keys each. Then we combine all these $mb+1$ keys, and split them into $m+1$ new nodes. Again, the same inequality (2) is needed.

Problem 5 PROBABILITY [15+10 Points]
Consider the following silly randomized algorithm: At each step, it rolls a dice (digital one, of course) that has one of six possible outcomes: $1,2,3,4,5,6$. In the i-th step, if the outcome is less than i, we stop. Otherwise, we go to the next step. The first step is $i=1$. For instance, we never stop after first step and surely stop by the 7-th step. Let T be the random variable corresponding to the number of steps.
(a) Set up the sample space, the event space, and the probability function for T.

ANSWER: Let Ω be the sample space and Σ the event space. The first mistake is to assume $\Omega = \{1,2,\ldots,6\}$. In our case, we can think of each sample point as the ENTIRE history of the rolls of your dice. Let $i_1, i_2, \ldots$ be the result of the first, second, etc, roll. For instance, if you stop after step 2, this means you rolled (i_1,i_2) where i_1 is unrestricted ($i_1 = 1,\ldots,6$) and $i_2 = 1$. If you stop after step 3, this means you rolled (i_1,i_2,i_3) where i_1 is unrestricted, $i_2 \geq 2$ and $i_3 \leq 2$. And so on. Then

$$\Omega = \{(i_1,i_2) : i_1 \geq 1, i_2 \leq 1\} \cup \{(i_1,i_2,i_3) : i_1 \geq 1, i_2 \geq 2, i_3 \leq 2\} \cup \{(i_1,i_2,i_3,i_4) : i_1 \geq 1, i_2 \geq 2, i_3 \geq 3, i_4 \leq 3\} \cup \ldots$$

Then $\Sigma = 2^\Omega$. So each $\omega \in \Omega$ is a finite sequence of length between 2 and 7. The probability of any sample point ω is

$$\Pr(\omega) = 6^{|\omega|}$$

where $|\omega|$ is the length of the sequence.

Now, you need to define T. This is a random variable, which means T is a function from Ω to real numbers:

$$T : \Omega \to \mathbb{R}.$$

In this case, it is very simple,

$$T(\omega) = |\omega|.$$

(b) Compute the expected value of T.

ANSWER: The expected value of T is

$$E[T] = \sum_{\omega \in \Omega} |\omega| \Pr(\omega).$$

It is a bit tedious to compute (and we would not take off many points if you make mistakes in your computation). But the correct answer is

$$E[T] = \frac{1223}{324} > 3.$$