CS G713: Advanced Algorithms	Problem Set 2
Assigned: October 10, 2008	Due: October 24, 2008 before class

Guidelines   If you write "I do not know how to solve this problem" then you get 1/4 of the score for the problem. If you write nonsense then you get 0.

As we are going to learn in this class, *time* and *space* are very valuable resources. Strive to give effective, compact solutions. Your solutions should touch on all the main points, but long calculations or discussions are not required nor sought.

Do not worry if you sometimes "do not get it." The problems are meant to stimulate you, not to overwork you. Unless specified otherwise, you can collaborate, but you must acknowledge all your collaborators in your solutions. To hand in your solutions: Give it to me, slide it under my door West Village H (246), or email it to csg713-instructor@ccs.neu.edu.

Problem 1. Universal hash functions without primes.   Your are to develop a universal hash function to hash elements from $\{0,1\}^u$ into $\{0,1\}^r$. However later you are to explain the construction to your Boss who never studied any number theory, so you cannot use the construction seen in class using prime numbers.

(1) Consider the following family of functions for a picked randomly in $\{0,1\}^u$: $h_a(x) := \sum_{i=1}^u a_i \cdot x_i$ modulo 2. Show that this is a universal family of hash functions for $r = 1$.

(2) Prove that if $h_a, h'_{a'}$ are universal families of hash functions, then the family (for random a, a') $(h_a, h'_{a'})(x) := (h_a(x), h'_{a'}(x))$ is also universal.

(3) Conclude that there is a solution to your hashing problem that uses randomness $u \cdot r$. Verify that this construction uses more randomness (and hence space) than the one seen in class. (You can ignore the fact that the construction seen in class requires the range to be a prime number, as this can be taken care of with only a slight error probability.)

(5) Now consider the following family based on convolution. For a random $a \in \{0,1\}^{r-1+u}$,

$$h_a(x) := (\sum_{i=1}^u a_{r-1+i} \cdot x_i, \sum_{i=1}^u a_{r-2+i} \cdot x_i, \ldots, \sum_{i=1}^u a_{1+i} \cdot x_i, \sum_{i=1}^u a_{0+i} \cdot x_i),$$

where the sum in each entry is modulo 2, and so the output of h is in $\{0,1\}^r$. Prove that this family is universal.

(6) Observe that the construction in (5) uses randomness close to the construction seen in class. (We may see later that this construction can be computed very efficiently.)

Problem 2. Expected height of a skip list.   Prove that the expected number of levels $E[h]$ in a random skip list of n elements is $E[h] = \theta(\log n)$. Note this requires both an upper and a lower bound.

Hint: For the upper bound first prove and then use the fact that any random variable X taking values in $\{0, 1, 2, \ldots\}$ satisfies $E[X] = \sum_{i \geq 1} \Pr[X \geq i]$. For the lower bound use Markov's inequality.

Problem 3. CLRS 17-2: Making binary search dynamic. Note: Recall from the mergesort algorithm that to merge two sorted lists we take linear time.

Problem 4. CLRS 17-3: Amortized weight-balanced trees. Notes: For part (a) you can use the simple fact that there is a linear time and linear space algorithm that given a binary search tree outputs all its keys in a sorted array. You may want to interpret part (d) as follows: Show that there exists a constant c, depending on α, such that whenever we rebuild a tree with m nodes, the potential is at least the cost to rebuild the tree (which is Am by part (a) for an absolute constant A).