Problem 1: In the discussion on hashing with chaining, it was assumed that the key to be searched is equally likely to be any of the $n$ inserted keys. Under this assumption, it was shown that the expected time for a successful search is $O(1 + \alpha)$, where $\alpha = n/m$ is the load factor. The proof also assumed that the insert function inserted each key at the beginning of a list. Show that the expected time for a successful search remains $O(1 + \alpha)$ even if the insert function inserts a key at the end of a list.

Solution to Problem 1: Suppose the key $k$ that is being searched for was the $i$th key inserted into the table. Note that $i - 1$ keys were inserted before key $k$. By the simple uniform hashing assumption, each list received $(i - 1)/m$ of these keys on the average. In other words, when things are inserted at the end, the expected position of key $k$ in its list is $1 + (i - 1)/m$. So, the expected number of keys examined to locate $k$ is $1 + (i - 1)/m$.

Since each key is equally likely to be searched, the average number of keys examined for locating any key is equal to $\frac{1}{n} \sum_{i=1}^{n} (1 + (i - 1)/m)$. With simple algebraic manipulations, this expression can be seen to be $1 + \alpha/2 - 1/(2m) < 1 + \alpha/2$. Adding the $O(1)$ time for hash value computation, we see that the expected time for a successful search is $O(2 + \alpha/2) = O(1 + \alpha)$.

Problem 2: Suppose we use a hash function $h$ to hash $n$ distinct keys into a hash table of size $m$. Assuming simple uniform hashing, compute the expected number of collisions.

Solution to Problem 2: For each pair of distinct keys $k_i$ and $k_j$, define an indicator RV $x_{ij}$, $1 \leq i < j \leq n$. Thus, $x_{ij} = 1$ if keys $k_i$ and $k_j$ collide and 0 otherwise. If $X$ denotes the RV that gives the total number of collisions, we have

$$X = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} x_{ij}.$$ 

We want to compute $E[X]$. By linearity of expectation, we have

$$E[X] = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} E[x_{ij}]. \quad (1)$$

We compute $E[x_{ij}] = \Pr\{x_{ij} = 1\}$ as follows. By the simple uniform hashing assumption, for any index $p$, the probability that both $k_i$ and $k_j$ hash to the value $p$ is equal to $1/m \times 1/m = 1/m^2$. Since $p$ can take on $m$ different values, the probability that $k_i$ and $k_j$ collide is $m \times 1/m^2 = 1/m$. In other words, $E[x_{ij}] = \Pr\{x_{ij} = 1\} = 1/m$. Since the number of indicator RVs is $n(n - 1)/2$, we have from Equation (1), the expected number of collisions $E[X] = n(n - 1)/(2m)$. 

---