Jade Yu Cheng
ICS 311
Homework 12
Oct 14, 2008

Question for lecture 15

Problem 12-3 on p. 250

Longest-probe bound for hashing
A hash table of size m is used to store n items, with $n \leq m/2$. Open addressing is used for collision resolution.

a. Assuming uniform hashing, show that for $i = 1,2,3, \dots, n$, the probability that the ith insertion requires strictly more than k probes is at most 2^{-k}.

Answer: In an unsuccessful search, every probe but the last accesses an occupied slot that does not contain the desired key, and the last slot probed is empty. Let us define the random variable X_k to be the number of probes made in an unsuccessful search, and let us also define the event A_k, for $k = 1,2,3, \dots$, to be the event that there is an kth probe and it is to an occupied slot. Then the event $\{X_k > k\}$ is the intersection of events $A_1 \cap A_2 \cap A_3 \cap \dots \cap A_k$. We will bound $Pr\{X_k > k\}$y bounding $\Pr\{ A_1 \cap A_2 \cap A_3 \cap \dots \cap A_k \}$.

$$\Pr\{ A_1 \cap A_2 \cap A_3 \cap \dots \cap A_k \}$$
$$= \Pr\{ A_1\} \cdot \Pr\{A_2 \mid A_1\} \cdot \Pr\{A_3 | A_1 \cap A_2\} \cdot \dots \cdot \Pr\{A_k \mid A_1 \cap A_2 \cap \dots \cap A_{k-1}\}$$

Since there are n elements and m slots, $Pr\{A_1\}/m$. For $j > 1$, the probability that there is a jth probe and it is to an occupied slot, given that the first $j - 1$ probes were to occupied slots, is $(n - j + 1)/(m - j + 1)$. This probability follows because we would be finding on of the remaining $(n - (j - 1))$ elements in one of the $(m - (j - 1))$ unexamined slots, and by the assumption of uniform hashing, the probability is the ratio of these quantities. Observing that n<m implies that $(n - j)(m - j) <= n/m$ for all j such that $0 <= j < m$, we have for all k such that $1 < k < m$,

$$
\begin{aligned}
\Pr\{\mathrm{X_k} \geq \mathrm{k}\} &= \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdot \frac{n-3}{m-3} \cdot \ldots \cdot \frac{n-(k-1)}{m-(k-1)} \\
&\leq \left(\frac{n}{m}\right)^k \\
&\leq \left(\frac{1}{2}\right)^k \\
&= 2^{-k}
\end{aligned}
$$

We consider the fact that we need more than kth probes in order to find the key. An alternative way to think about this situation is that the kth probe still returns with a collision. Therefore, the equation above indicates the probability of the ith insertion requires strictly more than k probes is at most 2^{-k}.

b. Show that for $i = 1,2,3, \ldots, n$, the probability that the kth insertion requires more $2lgn$ probes is at most $1/n^2$.

Answer: As discussed in the previous question, the probability that the ith insertion requires strictly more than k probes is at most 2^{-k}. We could plug in the value in this specific case for this problem. The probability that the ith insertion requires more than $2lgn$ probes is at most $1/n^2$.

$$
\begin{aligned}
\Pr\{\mathrm{X_{2lgn}} \geq \mathrm{k}\} &= \frac{n}{m} \cdot \frac{n-1}{m-1} \cdot \frac{n-2}{m-2} \cdot \frac{n-3}{m-3} \cdot \ldots \cdot \frac{n-(2\lg n-1)}{m-(2\lg n-1)} \\
&\leq \left(\frac{n}{m}\right)^{2\lg n} \\
&\leq \left(\frac{1}{2}\right)^{2\lg n} \\
&= 2^{-2\lg n} \\
&= \frac{1}{n^2}
\end{aligned}
$$

Let the random variable X_i denote the number of probes required by the ith insertion. You have shown in part (b) that $\Pr\{X_k \geq k\} \leq 1/n^2$. Let the random variable $X = \max_{1 \leq i \leq n} X_i$ denote the maximum number of probes required by any of the n insertions.

c. Show that $\Pr\{X \geq 2\lg n\} \leq 1/n$.

Answer: We discussed for each individual time of insertion. We calculated the probability of strictly more than $2lgn$ probes for the ith time of insertion. This problem is to compute the overall probability. That is the probability of the maximum number of probes required by any of the n times of insertions. This is simply an individual events series, and the probability of the overall probability is simply the summation of all of the events. If any one of these events happens, the overall event happens. To be more specific, if any of the n times of insertions happens to require a number of $2lgn$ probes, we say a probe with length $2lgn$ happens within the n times of insertion.

$$
\begin{aligned}
&\Pr\{X_1 \geq 2\lg n\} + \Pr\{X_2 \geq 2\lg n\} + \Pr\{X_3 \geq 2\lg n\} + \ldots + \Pr\{X_n \geq 2\lg n\} \\
&= \sum_{1 \leq i \leq n}^{i} \Pr\{X_i \geq 2\lg n\} \\
&\leq \sum_{1 \leq i \leq n}^{i} \frac{1}{n^2} \\
&= \frac{1}{n^2} \cdot n \\
&= \frac{1}{n}
\end{aligned}
$$