

Family of uniform hash functions

The notion of pairwise independence says that, for any $x_1 \neq x_2$ and $c_1, c_2 \in Z_p$, we have that

$$\Pr_{h \in \mathcal{H}} [h(x_1) = c_1 \wedge h(x_2) = c_2] = \Pr_{h \in \mathcal{H}} [h(x_1) = c_1] \times \Pr_{h \in \mathcal{H}} [h(x_2) = c_2]$$

In other words, the joint probability is the product of the two individual probabilities. Show that the family of hash functions $H = \{h_{ab}(x) = ((ax + b) \bmod p) \bmod m : a \in Z_p^*, b \in Z_p\}$ (seen in class) is "pairwise independent", where p is a sufficiently large prime number ($m + 1 \leq p \leq 2m$).

SOLUTION

Each probability from the right hand side of the equation are equal to $\frac{1}{m}$, simply because there are m buckets chosen uniformly (Universal hash function). Therefore, in the right hand side we have a probability equal to $\frac{1}{m^2}$.

On the left side of the equation, we will first count all the possible solutions (x_1, x_2) , and then we will count only the good solutions (that is, the solutions x_1 and x_2 that match, respectively, c_1 and c_2).

$h(x_1)$ and $h(x_2)$ can be rewrite with their expression, such as:

$$\left. \begin{aligned} h(x_1) &= (a \cdot x_1 + b \bmod p) \bmod m \\ h(x_2) &= (a \cdot x_2 + b \bmod p) \bmod m \end{aligned} \right\} x_1 \neq x_2$$

Now, let's consider the inner modulo p expressions as $r = a \cdot x_1 + b \bmod p$ and $s = a \cdot x_2 + b \bmod p$, since p is prime and there is a linear transformation). Thus we can rewrite the original expression as: $\Pr[r \bmod m = c_1 \wedge s \bmod m = c_2]$. Since $a \in [1 \dots p-1] = Z_p^*$ and $b \in [0 \dots p-1] = Z_p$, and both a and b are fixed for r and s , then there are $p \cdot (p-1)$ possible different combination for x_1 and x_2 ($x_1 \neq x_2$).

Now, we focus on the "good" solutions. When p is sufficiently larger than m , the number of x_i that will match a given value c_i are $\frac{p}{m}$, as we have seen in class. However, we need to take in account modular arithmetic: not every values of m is going to be a multiple of p . Adding this to the analysis, the number of matches are bound by considering the round-to-floor and round-to-ceil limits:

$$\left\lfloor \frac{p}{m} \right\rfloor \leq \# \text{ OF } x_i \text{ MATCHING } c_i \leq \left\lceil \frac{p}{m} \right\rceil$$

Since we have both c_1 and c_2 , we need to consider their joint probability (i.g. multiplication), and then we can be bound it by combining both pieces of the analysis (dividing "good" solutions by all possible combination), giving us an approximation to the expected $\frac{1}{m^2}$:

$$\frac{1}{p \cdot (p-1)} \left(\left\lfloor \frac{p}{m} \right\rfloor \right)^2 \leq \Pr[r \bmod m = c_1 \wedge s \bmod m = c_2] \leq \frac{1}{p \cdot (p-1)} \left(\left\lceil \frac{p}{m} \right\rceil \right)^2$$

It must be noted that the $\frac{1}{m^2}$ result may not be fully reached since that all the possible values a are $p-1$, which make it impossible to remove the $\frac{p}{p-1}$ term. The only possibility is when $a = 0$ that would not make any sense algorithmically speaking. Therefore the result can only be approximated.