

Notes on Universal Hash Functions, Part 1

We proved in Theorems 11.1 and 11.3 that if we take n items and insert them into random locations in a hash table with m addresses, and we consider any particular item x , the expected number of other items y which hash to the same address - that is, which collide with x in the hash table - is n/m .

This proof is based on the idea that x and y are hashed independently into the table, so that $\Pr[h(x) = h(y)] = 1/m$. Unfortunately, implementing a hash function by hashing items to random addresses is not really feasible.

But if we can come up with a hash function which is easy to implement, and still has the property that $\Pr[h(x) = h(y)] = 1/m$, then the proof we used in the random case goes through and we still can guarantee that the expected number of elements that collide with x is n/m . Such a hash function is called *universal*.

We have the following example of a universal hash function for integer items. We use a prime number p which is larger than our largest possible item. At run-time, we pick two random integers,

$$a \in \{1, \dots, p-1\}$$

and

$$b \in \{0, \dots, p-1\}$$

Then our hash function is:

$$h(x) = ((ax + b) \bmod p) \bmod m$$

Obviously modular arithmetic is very important in analyzing this hash function. We'll prove one handy modular arithmetic fact.

Claim 1 *Say p is prime and we have $a, x, y \in \{0, \dots, p-1\}$. Then if*

$$ax \bmod p = ay \bmod p$$

either $x=y$, or $a=0$.

Proof: Assume without loss of generality that $x \geq y$. We have

$$(ax - ay) \bmod p = a(x - y) \bmod p = 0$$

So there are two possibilities: either $a(x - y) = 0$, in which case the Claim is true, or p divides $a(x - y)$. But both a and $(x - y)$ are numbers smaller than p . Each can be expressed uniquely as a product of primes, all of which are smaller than p . So $a(x - y)$ can be expressed as a product of primes less than p , and since the representation of a number as a product of primes is unique, p does not divide $a(x - y)$. That leaves only the possibility that $a(x - y) = 0$, in which case the Claim is true. \square

Armed with this Claim, let us prove that $h(x)$ is universal. Let us begin by studying the the first part of the function, $(ax + b) \bmod p$. Let's call this $g(x)$.