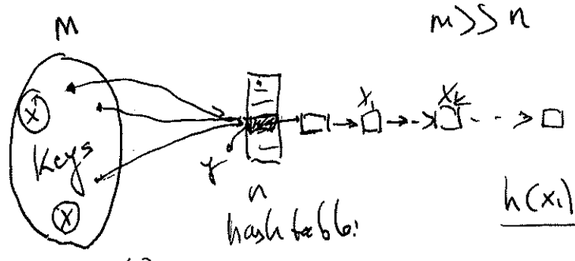
Lecture 7

Last Time: hash functions



$h(x) \to y$

$h(x_1) = h(x_2) = y$. hash collision!

Universal hash functions $\boxed{P(h(x) = h(x')) = \frac{1}{n}}$ $\Rightarrow$ Randomization.

Approach 1: $H = \{ h_{a,b}(x) = ((ax+b) \bmod m) \bmod n \mid a, b \in \mathbb{Z}_m, a \neq 0 \}$

n can be any integer but m should be a prime number

pick one hash function at random from H and use it.

Today: why is this a good hash function (i.e. universal hash fn.)

to show this is a universal hash function we need to show:

$$P(h_{a,b}(x) = h_{a,b}(y)) = \frac{1}{n}$$

$h_{a,b}(x) = ((ax+b) \bmod m) \bmod n \Rightarrow P(r \bmod n \equiv s \bmod n)$

$h_{a,b}(y) = ((ay+b) \bmod m) \bmod n$

Let's fix r and consider how many s values can be congruent to r mod n.

why. because

$$P(r \equiv s \bmod n) = \frac{\#\ s \text{ values with } r \equiv s \bmod n}{\#\ \text{of possible } s \text{ values}} = \frac{\lceil \frac{m}{n} \rceil - 1}{|\{0, \ldots, m-1\}|} \leq \frac{\left(\frac{m+n-1}{n}\right) - 1}{n-1}$$

$= m - 1$

$$\leq \frac{\frac{m}{n} + \frac{n}{n} - \frac{1}{n} - 1}{n-1} \leq \frac{\frac{n-1}{n}}{n-1} = \frac{1}{n}$$

∎

Probabilistic Analysis of an Algorithm

1. Exact (more detailed and challenging)
2. fast and good approximation by using Indicator Random Variables
   takes value either 1 or 0

Example: Birthday Paradox

$n = 365$ days

how many people (k) must there be in a room before 50% chance that 2 of them are born on the same day of the year.

$\Rightarrow$ 50% chance

$\Rightarrow$ probabilistic analysis

- let's assign each person in the room an index. $i = 1 \dots k$.
- Assume the birthdays (b_i) are independent $\quad 1 \le b_i \le n$

$$P(b_i = r) = \frac{1}{n} \qquad r = 1 \dots n, \; i = 1 \dots k$$

$$P(b_i = r \text{ and } b_j = r) = \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n^2}$$

$$P(b_i = b_j) = \sum_{r=1}^{n} P(b_i = r \text{ and } b_j = r) = \sum_{r=1}^{n} \left(\frac{1}{n^2}\right) = \frac{1}{n}$$

Continue with Indicator Random Variables

define a random variable r.v. $X_{ij} = I\{\text{person } i \text{ and person } j \text{ have the same birthday}\}$

$$X_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ have the same bday} \\ 0 & \text{o.w.} \end{cases}$$

$$X = \sum_{i=1}^{k} \sum_{j=i+1}^{k} X_{ij}$$

$$\text{Prob}(\text{EVENT}) = \frac{\#\text{ of times that event occurs}}{\#\text{ of times all the events occur}} \quad \text{(total)}$$

Approach 2: how to represent x in some base n

$m = 1000$
$n = 11$

$x = 999 \Rightarrow$ (implicitly in base 10)

$x = 999 = 8 \cdot 11^2 + 2 \cdot 11^1 + 9 \cdot 11^0 \Rightarrow (8,2,9)_{11}$

$$H(x) = \left\{ (a_1 x_1 + a_2 x_2 + \cdots + a_k x_k) \bmod n \;\middle|\; (a_1, a_2 \cdots a_k) \in \mathbb{Z}_n^k \right\}$$

($a_1, a_2 \cdots a_k$: randomly chosen)

$$H(x) = \left(\sum_{i=1}^{k} a_i x_i \right) \bmod n$$

ex! $a_1 = 9 \quad a_2 = 1 \quad a_3 = 2$

$H(999) = (8 \cdot 9 + 2 \cdot 1 + 9 \cdot 2) \bmod 11 = 4$

why is this a universal hash function?

$$P(h(x) = h(y)) = \frac{1}{n}$$

(read the proof in pg 45 of the text book)

Now consider a key $\boxed{x}$ and compute EXPECTED lookup time

$$Y_i = \begin{cases} 0 & \text{o.w.} \\ 1 & \text{if } x \text{ and } y_i \text{ has the same index} \end{cases}$$

$$E\left[\sum_{i=1}^{n} Y_i \right] = \sum_{i=1}^{n} E[Y_i] = \sum_{i=1}^{n} \frac{1}{n} = 1$$

(collisions)

due to "Linearity of expectation"

$$E[X] = E\left[\sum\sum x_{ij}\right] = \sum\sum E[x_{ij}] = \sum\sum P[x_{ij}] = \binom{k}{2}\frac{1}{n}$$ (34)

linearity of expectation

$\frac{1}{n}$.

$$= \frac{k(k-1)}{2n} \Rightarrow k(k-1) \geq 2n. \Rightarrow k = \sqrt{2n}+1$$

$k^2 - k \geq 2n$

$\sqrt{k^2+k} \geq \sqrt{2n}$.

So if we have $k = \sqrt{2n}+1$ people in the room we expect at least 2 to have the same birth day

$\sqrt{k^2-k} \geq 1.4\sqrt{n}$ $k \geq \sqrt{n}$

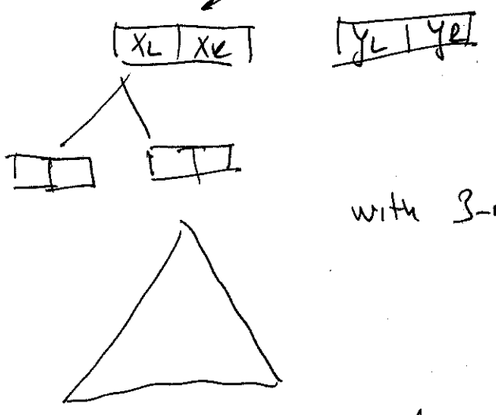
$k = \sqrt{2 \cdot 365}+1 \Rightarrow k \cong 28$

Divide and Conquer Algorithms

ex: multp(3) (x, y)

$xy = \left(2^{n/2}x_L + x_R\right)\left(2^{n/2}y_L + y_R\right)$

$\Rightarrow$ 4 multiplications



$$T(n) = 4\,T\left(\frac{n}{2}\right) + \underbrace{\theta(n)}_{\text{book keeping}}$$

$\Rightarrow \theta(n^2)$

with 3-multiplications

$$T(n) = 3\,T\left(\frac{n}{2}\right) + \theta(n)$$

$\Rightarrow \theta(n^{1.595})$

we have several parameters
1. problem size n
2. # of sub problems a
3. size of each subproblem $\frac{n}{b}$
4. how much it costs to combine the answers

Theorem

If $T(n) = a \cdot T\left(\left\lceil \frac{n}{b} \right\rceil\right) + \theta(n^d)$

for some constants $a > 0$ $b > 1$ $d \geq 0$

$$T(n) = \begin{cases} \theta(n^d) & \text{if } d > \log_b a \quad \text{case I} \\ \theta(n^d \lg n) & \text{if } d = \log_b a \quad \text{case II} \\ \theta(n^{\log_b a}) & \text{if } d < \log_b a \quad \text{case III} \end{cases}$$

case I, II, III are based on the behavior of geometric series.

decreasing — fixed — increasing

$$T(n) = 2T(n/2) + \theta(n) \quad \text{how to solve this?}$$

cn

example:

① way

$$T(n) \leq 2T\left[2T\left(\frac{n}{4}\right) + \frac{cn}{2}\right] + cn.$$

$$\leq 4\left[2T(n/8) + \frac{cn}{4}\right] + 2c$$

$$\leq \cdots$$

$$T(n) \leq 2^k T\left(\frac{n}{2^k}\right) + kcn.$$

substitute $k = \lg n$

$$T(n) \leq n\,T(1) + cn \lg_2 n \Rightarrow O(n \lg n).$$

② Apply master 定理 theorem

next we will prove the theorem.