April 19

HW #5 on Learn due May 3

Recall.

Let $a, b \in \mathbb{Z}^+$, $a \geq b$

$a = qb + r \quad (0 \leq r < b)$

(1) if $r = 0$, then $\gcd(a,b) = b$

(2) if $r \neq 0$, then $\gcd(a,b) = \gcd(b,r)$

Recursion.

$a = q_1 b + r_1$

$b = q_2 r_1 + r_2$

$r_1 = q_3 r_2 + r_3$

$\vdots$

Running Time

$\lg q_1 \cdot \lg b = \lg(a/b) \lg b$

$= (\lg a - \lg b) \lg b$

$\lg q_2 \cdot \lg r_1 = (\lg b - \lg r_1)(\lg r_1)$

$\lg q_3 \cdot \lg r_2 = (\lg r_1 - \lg r_2) \lg r_2$

$\vdots$

Running Time Analysis:

input size: number of bits. $\log_2 a + \log_2 b$

running time:

$a = 10\phi\phi1001 \quad b = 1001$

$$
\begin{array}{r}
10\phi\phi1001 \\
1001 \\
\hline
101001 \\
1001 \\
\hline
1001 \\
1001
\end{array}
$$

number of subtraction $\lg \frac{a}{b}$

each subtraction $\lg b$.

Running time:

$$\begin{aligned}
&(\lg a - \lg b)\lg b + && &&(\lg a - \lg b)\lg b + \\
&(\lg b - \lg r_1)\underline{\lg r_1} + && \leq &&(\lg b - \lg r_1)\lg b + \\
&(\lg r_1 - \lg r_2)\underline{\boxed{\lg r_2}} + && &&(\lg r_1 - \lg r_2)\underline{\boxed{\lg b}} + \\
&\quad \vdots && &&\quad \vdots
\end{aligned}$$

$$\leq \underline{\lg a \cdot \lg b}$$

<u>This running time is amortized</u>, in contrast to:

$$(\lg a)\left((\lg a - \lg b)\lg b\right) = \underline{\underline{\lg^2 a \lg b}}$$

Recall:

$U = \{0, 1, \cdots, M-1\}$ we have a table of size n. ($n \ll M$)

Find a prime $p \geq M$

Generate two random numbers. $a \sim \{1, 2, \cdots, p-1\}$

$b \sim \{0, 1, \cdots, p-1\}$

hash function $h_{a,b}(x) = ((ax+b) \bmod p) \bmod n$

The family $H = \{ h_{a,b} \mid a \in \{1, \cdots, p-1\}, b \in \{0, 1, \cdots, p-1\} \}$

is universal.

For any pair of distinct keys x, y, when we randomly pick a hash function h from H, the odds of collision is $\leq \frac{1}{n}$

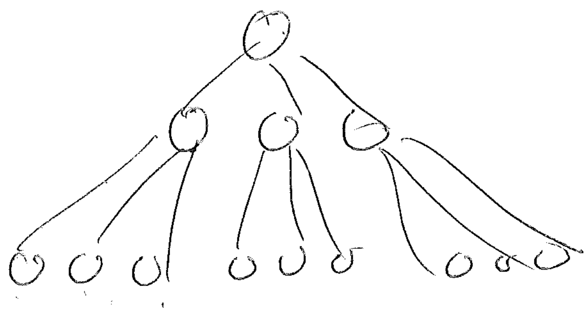
Assume we are given m keys, whats the expected number of collisions per hash table entry

$M = 10^9$

$m = 30{,}000$

$n = 10{,}000$

$\frac{m}{n}$



$$\log_3 n = \frac{\log_2 n}{\log_2 3}$$

$H = \{ h_{a,b}(x) = (ax+b) \% p \% n \mid a \in \{1, 2, \cdots, p-1\},\ b \in \{0, 1, 2, \cdots p-1\} \}$

$x \neq y$, two distince keys

① $s = (ax+b) \bmod p$ $\quad$ $t = (ay+b) \bmod p$

entry in the table

$s \bmod n$ $\quad$ $t \bmod n$

② ① for $x \neq y$, ~~then~~ $s \neq t$

thus whether x, y collide depend on $s \bmod n \stackrel{?}{=} t \bmod n$

③ ② for $h_{a_1,b_1} \neq h_{a_2,b_2}$

$(s_1, t_1) \neq (s_2, t_2)$

there is a one-to-one correspondence between (s, t) pairs to $h_{a,b}$

③ the number of hash funceons causy collision is the number of (s, t) pairs with $s \bmod n = t \bmod n$

if $x \neq y$ then $s \neq t$

$$\boxed{\begin{aligned}(ax+b) \bmod p &= s \\ (ay+b) \bmod p &= t\end{aligned}}$$

Pf. Assume Not, then $s = t$

$(ax+b) \bmod p = s$ (1)

$(ay+b) \bmod p = t = s$ (2)

(1) − (2) $(ax+b)\%p - (ay+b)\%p = 0$

$((ax+b) - (ay+b))\%p = 0$

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

$0 \le x \le p-1$

$0 \le y \le p-1$

Thus $p \mid a(x-y)$

$-(p-1) \le x-y \le p-1$
and $x-y \neq 0$

However $p \nmid a$ $p \nmid (x-y)$ p is prime

so $p \nmid a(x-y)$ a contradiction!

the assumption is wrong! and $s \neq t$.

$$\left|\{(s,t) \mid s \neq t\}\right| = \text{\# of 2 permutations on } p \text{ elements} = \frac{p!}{(p-2)!} = \underline{\underline{p(p-1)}}$$

$s, t, \in \{0, 1, \cdots, p-1\}$

$$_nP_r = \frac{n!}{(n-r)!}$$

$$\left| H = \{h_{a,b}\} \right| = \underline{\underline{p(p-1)}}$$

(3) Assume we have already shown there is a one-to-one correspondence between $\{(s,t) \mid s \neq t\}$ to $\{h_{a,b}\}$

And from (1) whether ~~s,t~~ x, y collide depend on whether $s \bmod n$ equal to $t \bmod n$

Thus, the number of hash functions causing x, y collide is the number of (s,t) such that $s \bmod n = t \bmod n$

$\{(s,t) \mid s \neq t \ , \ s,t \in \{0,1,\cdots,p-1\}\}$

How many (s,t) pairs will ~~give~~ or have $s \equiv t \bmod n$?

Eg. $\{(s,t) \mid s \neq t, \ s,t \in \{0,1,\cdots,6\} \ p=7\}$

$n=3$ $\qquad 7 = 2\times 3 + 1$

mod3 = 0	mod3 = 1	mod3 = 2
0	1	2
3	4	5
6.		

$$\frac{6+2+2}{7\times 6} = \frac{10}{7\times 6} = \frac{10}{42} < \frac{1}{3}$$

$0 \sim p-1$

$\bmod n=0$ $\quad$ $\bmod n=1$ $\quad$ $\bmod n=2$ $\quad \cdots \quad$ $\bmod n=n-1$

$\approx \frac{p}{n}$

$$\frac{\cancel{n}\cdot\frac{p}{\cancel{n}}\left(\frac{p}{n}-1\right)}{\cancel{p}(p-1)} = \frac{\cancel{\frac{p-n}{n}}}{\cancel{p-1}} = \frac{1}{n}$$

$p = q_i n + r$

$\bmod n = 0$ $\quad$ $\bmod n = 1$ $\quad$ $\bmod n = n-1$

② We need to establish a one-to-one correspondence between $\{(s,t) \mid s \neq t\}$ to $\{h_{a,b}\}$

it suffices to show that

~~fors~~ for $(a_1,b_1) \neq (a_2,b_2)$

$(s_1,t_1) \neq (s_2,t_2)$

Proof. By Contradiction,

Assume Not, ~~thus~~ for $(a_1,b_1) \neq (a_2,b_2)$

$(s_1,t_1) = (s_2,t_2)$

There are three scenarios for $(a_1,b_1) \neq (a_2,b_2)$

① $a_1 \neq a_2$, $b_1 = b_2$

$s_1 = (a_1 x + b_1) \bmod p$ (a) $=$ $s_2 = (a_2 x + b_2) \bmod p$ (b)

$t_1 = (a_1 y + b_1) \bmod p$ (c) $=$ $t_2 = (a_2 y + b_2) \bmod p$ (d)

(a)$-$(b) $\left((a_1 - a_2)x + (b_1 - b_2)\right) \bmod p = 0$. (e)

(c)$-$(d) $\left((a_1 - a_2)y + (b_1 - b_2)\right) \bmod p = 0$ (f)

(e)$-$(f) $\left((a_1 - a_2)(x-y)\right) \bmod p = 0$ impossible

Scenario 2 $a_1 = a_2$ $b_1 \neq b_2$

Similarly

Scenario 0 $a_1 \neq a_2$ $b_1 \neq b_2$

Similarly.

String Matching.

Problem Given a string $s[1..n]$ and pattern $p[1..m]$, $m<n$ over some alphabet Σ, determine if p occurs in s.

For $j = 1$ to $n-m+1$

compare p to $s[j..j+m-1]$

Running time $O(nm) \longrightarrow O(n+m) \longrightarrow O(n)$