Hashing 2:

Feb 18 2021

"Two-level hashing"

First level: table size $O(N)$

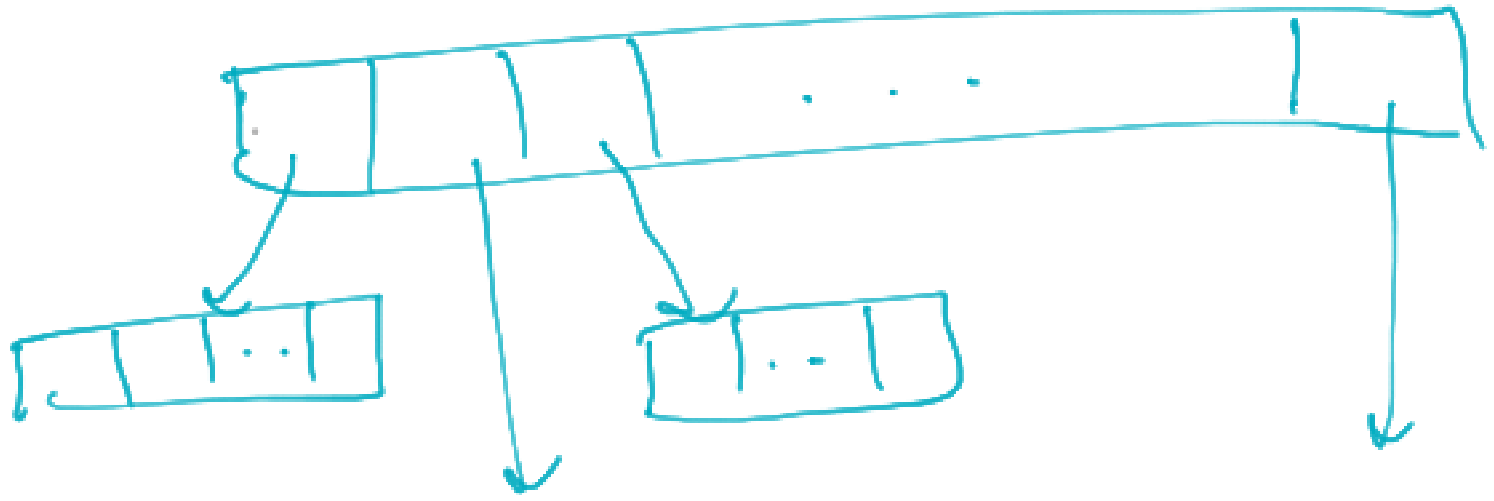
At each location: hash table (collision-free)

M = table size

$N = |S|$

Let

i = location index in the first level table.



$c(i)$ = num. of elements mapped to location i in the first level table

Q: What should be the size of table at i for collision free?

$c(i)^2$

Total table size:

$$\sum_{i=1}^{M} C(i)^2$$

C = total num. of collisions

$$C = \sum_{i=1}^{M} \binom{C(i)}{2}$$

Recall $E[C] = \binom{N}{2} \cdot \frac{1}{M}$

$$\Rightarrow \quad E\left[\sum_{i=1}^{M} \binom{C(i)}{2}\right] = \binom{N}{2} \frac{1}{M}$$

$$E\left[\sum_{i=1}^{M} C(i)^2 - \sum_{i=1}^{M} C(i)\right] = O(N)$$ since $M = O(N)$

$$E\left[\sum_{i=1}^{M} C(i)^2\right] = O(N) \qquad \because E\left[\sum_{i=1}^{M} C(i)\right] = O(N)$$

Total table size = $O(N)$

Collision-free in $O(N)$ space

"Perfect hashing"

---

k-wise independent hash functions:

Defn: A family $\mathcal{H}$ of hash functions $U \to [M]$ is k-wise independent if for any k distinct keys $x_1 \dots x_k$ and any k values $\alpha_1, \dots, \alpha_k$ we have

$$P\left(h(x_1) = \alpha_1 \cap h(x_2) = \alpha_2 \cap \dots h(x_k) = \alpha_k\right) \le \frac{1}{M^k}$$

For $k = 2$, "pairwise independent".

Suppose $\mathcal{H}$ is k-wise indep. $(k \geq 2)$

① $(k-1)$ wise indep

② for any $x \in U$ and value $a \in [M]$

$$P[h(x) = a] \leq \frac{1}{M}$$

③ Universal

Q: Which is stronger : pairwise indep. or universal ?

E.g. Construction from last class A

$$h(x) = Ax$$

$$h(0) = 0$$

$$P(h(0) = 0] = 1$$

Construction 1: 2-wise independent

$$h(x) = \underset{(m \times u)}{A}x + \underset{(m \times 1)}{b} \longrightarrow \text{m-length random binary vector.}$$

(modulo 2)

<u>Claim</u>: ...is 2-wise independent

Q: num. of hash fns? 2^{um+m}

Q: bits to store? $O(um)$

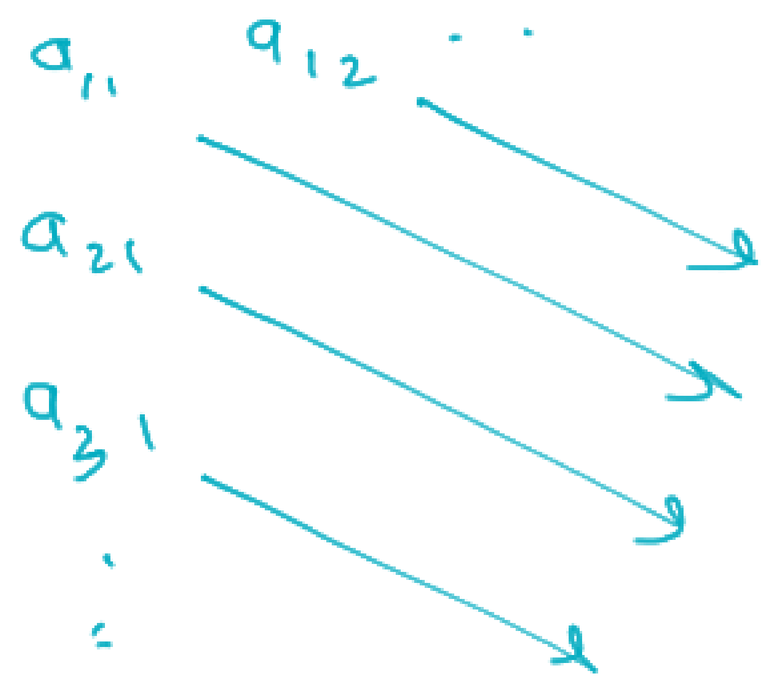
Can we do this with fewer bits?

Construction 2 (using fewer bits)

- fill first row & column with uniform random binary entries

- $A_{i,j} = A_{i-1,j-1}$

- $h(x) = Ax + b \pmod 2$



Claim: ... 2-wise indep.