Feb 16 2021

Hashing (Cont. from slides)

Universal Hash functions :

Due to Carter & Wegman (1979)

Defn : family H maps $U \to [m]$
is universal if for any $x \neq y \in U$

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

$h \in H$

Construction:

$|U| = 2^u$ Let $\underset{m \times u}{A}$ ← random binary entries

$|M| = 2^m$

For any $x \in U$ [u-length binary vector]

$$h(x) := Ax \pmod 2$$

Q: How many hash fns in the family? 2^{um}

Thm: .... is universal

Proof: $h(x) = h(y)$ for $x \neq y$

$Ax = Ay$

$A(x-y) = 0$

$Az = 0$ for $z \neq 0$

$y \because x \neq y$

we want to show $P(Az=0) \leq \frac{1}{M}$ for any $z \neq 0$

Let $z_{i^*} \neq 0$ $\exists i^*$ since $z \neq 0$

$Az = \sum A_j z_j$

column of A

$$Az = 0$$

$$\sum A_j z_j = 0$$

$$A_{i^*} = -\sum_{j \neq i^*} A_j z_j$$

$$\begin{bmatrix} A_{i^*1} \\ \vdots \\ A_{i^*m} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}$$

fixed vector of size m

m length binary vector (random)

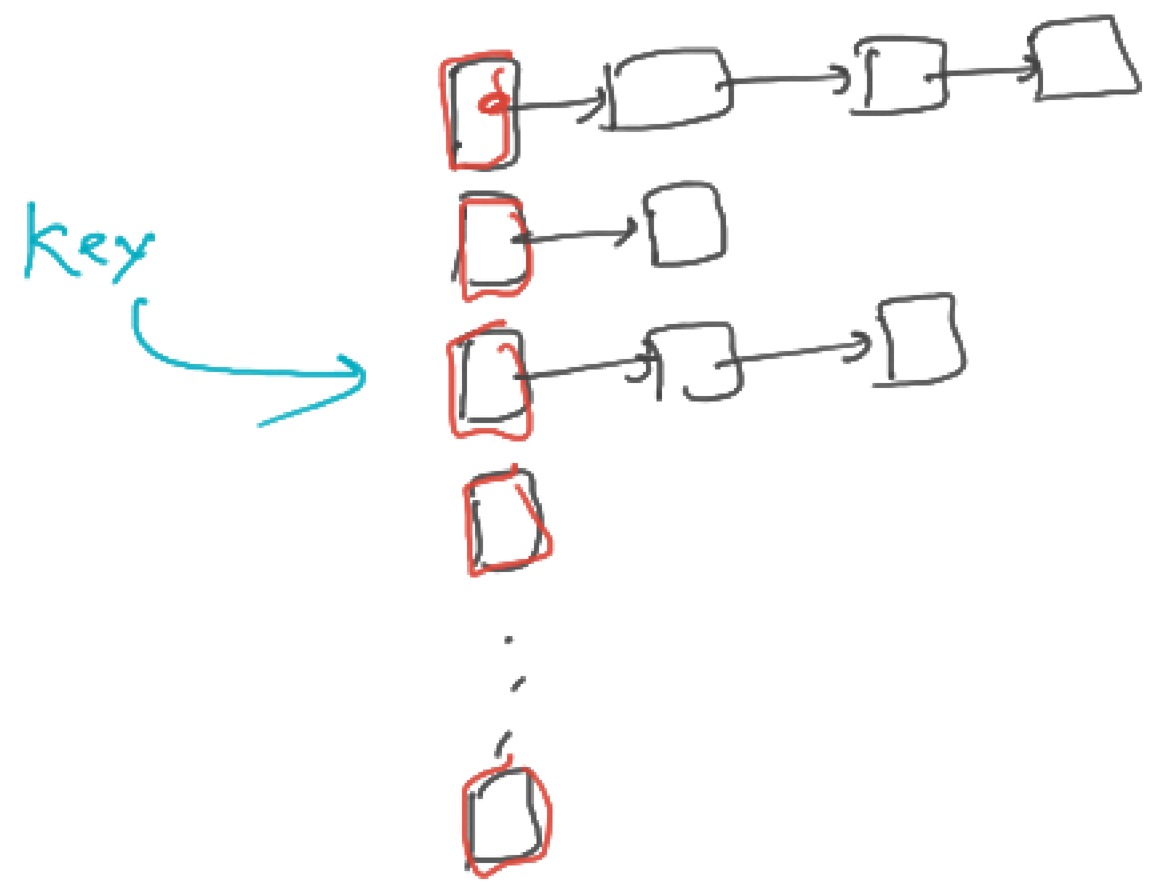
$$\text{Prob. of above} = \left(\frac{1}{2}\right)^m = \frac{1}{2^m} = \frac{1}{M}$$

Application: Hash table

Handling Collisions:

Approach 1: closed addressing

aka separate chaining.



Look up time $\approx$ Length of the list $\approx$ number of collisions

C_x = num. of elements mapped to the same value where x is mapped to.

L_x = len. of linked list containing x

$L_x = C_x + 1$

Q: What is $E[L_x]$?

$E[L_x] = 1 + E[C_x] = 1 + \frac{(N-1)}{M}$

If we choose $M \geq N$

$E[L_x] \leq 2$

Look up time constant in expectation!

<u>Recall</u>:

$|S| = N$

mapping to $[M]$

M = table size

C = total number of collisions

Q: $E[C]$?

$$\leq \binom{N}{2} \frac{1}{M}$$

Suppose $M \geq N^2 \Rightarrow E[C] \leq \frac{1}{2}$

Prob. [there exists a collision] = ?

$\frac{1}{2}$

Constant time look up (even in worst case).

But we need $M \geq N^2$

Can we get this with $O(N)$