Universal and Perfect Hashing

Hashing basics

- keys from universe $U$
- set $S \subseteq U$ of keys we care about
  \[ |S| = N \]
- hash table $A$ of size $M = \{0, 1, \ldots, M-1\}$
- hash function $h: U \rightarrow CM$
- want to store $x \in U$ in $A[h(x)]$
  what if $h(x) = h(y)$?

- Collision resolution: separate chaining
**Insert** \( (x) \):

- *Add to beginning of list at \( AH(x) \)*

**Lookup** \( (x) \):

- walk down list at \( AH(x) \) until find \( x \)

**Delete** \( (x) \):

- walk down list at \( AH(x) \) until find \( x \), delete it.

Q: What should we use as \( h \)?

Properties we want:

- Few collisions
- Small \( M \) (\( M = O(N) \))
- \( h \) fast to compute, easy to store
Theorem: For any hash function \( h \), if 
\[ |U| \geq (N-1)M + 1, \]
then there exists set \( S \) of \( N \) elements that all hash to same location.

Proof: Pigeonhole.

By Pigeonhole, \( J \in \{ M \} \) s.t.
\[ \# \text{elts of } U \text{ hash to } i \geq \frac{|U|}{M} > N-1 \]

Then

Random Function: When first sees
choose \( h(x) \) u.a.r. from \( CM \)
Def: A probability distribution \( H \) over hash fn's \( h : \mathcal{U} \rightarrow [M] \) is universal if
\[
\Pr_{h \sim H} \left[ h(x) = h(y) \right] \leq \frac{1}{m}
\]
for all \( x, y \in \mathcal{U}, x \neq y \).

Thm: If \( H \) is universal, then \( \forall S \subseteq \mathcal{U} \)
\[
|S| = N, \quad \forall x \in \mathcal{U}, \text{ expected}
\]
\( \text{hash collisions} \) by \( x \), elts of \( S \) is
\[
\leq \frac{N}{m}
\]
\( \exists \forall y \) in expected \( O(\sqrt{N/m}) \) time.

Ex: let \( c_{xy} = \begin{cases} 1 & x = y \lor h(x) = h(y) \\ 0 & \text{else} \end{cases} \)
\[
\mathbb{E} \left[ \mathbb{E} \left[ c_{xy} \right] \right] = \mathbb{E} \left[ \mathbb{E} \left[ c_{xy} \right] \right] \leq \frac{1}{m}
\]
\[
\geq \left( \text{time } L \text{ ops} \right) = O \left( \sqrt{N} \cdot \frac{m}{N} \right)
\]
Def: If \( H \) is universal and is uniform \( \Pr \) over \( \{h_1, h_2, \ldots, h_M\} \), then we say \( H \) is a universal hash family.

\[ u \in \{0,1\}^u \]

\[ M = 2^b \Rightarrow \text{All indices have } b \text{ bits} \]

\( H \) is all binary \( b \times n \) matrices

\[
\begin{align*}
S & \begin{pmatrix}
1 & 0 & \cdots & 0 \\
\vdots & \ddots & \ddots & \vdots \\
0 & \cdots & 1 & 0
\end{pmatrix} \\
& \times \begin{pmatrix}
0 & 1 \end{pmatrix} = \begin{pmatrix}
0 \\
1
\end{pmatrix}
\end{align*}
\]

\( h \) \( \text{hit} \)

\( h(x) = x \cdot h \)

Then: \( \Pr \left[ \text{hit}(x) = h(y) \right] = \frac{1}{M} \)

\( h \sim H \)

\( \forall x, y \in U \subseteq \{0,1\}^n \)

\( x \neq y \)
First choose all cols of \( h \) except col \( i \).

\( h(x) \) determined \( h(y) = h(x) \) if \( \forall \) th col s.t. \( h(x) - \text{current } h(y) \).

\( \Pr[\text{t} h(y) = h(x)] = \frac{1}{2^b} = \frac{1}{M} \)

**Perfect Hashing**

Assume \( kn \leq S \), new changes

**Method 1:** \( O(N^2) \) space.

Thus, left \( h \) universal with \( M = N^2 \).

Then \( \Pr[\text{no collisions}] \geq \frac{1}{2} \)
There are \( \binom{N}{2} \) possible collisions.

Each happens with \( \leq \frac{1}{M} = \frac{1}{N^2} \).

So if there is a collision \( \leq \binom{N}{2} \cdot \frac{1}{N^2} = \frac{N(N-1)}{2N^2} < \frac{1}{2} \).

---

**Method 2**: \( O(N) \) space

\( M = N \) universal \( u = \text{CM} \)

Draw \( h \sim u \)

- Let \( n_i = \# \text{clits in } S \) hash to \( i \)
For every $i$

**Build table of size $c_i^2$ at $A(c_i)$**

Choose $h_i$ universal for $[h_i \rightarrow c_i^2]$

Use perfect hashing with $0(c_i^2)$ space for $c_i^2$ hash families.

$$s_i = N + \sum_{i=1}^{N} c_i^2$$

Then $\Pr \left[ \sum_{i=1}^{N} c_i^2 > 4N \right] < \frac{1}{2}$