Resolving collisions w/ open addressing assuming \( n \leq m \)

The point is to avoid auxiliary linked lists. Use that space for table. Instead, create a probe sequence as a function of key value.

\( \text{permutation of slots to try.} \)

ex: \( h(64) \rightarrow 9, 2, 4, 8, 1, 3, 11, 7, 10, 5, 6. \)

Insert(64): Try \( T[9] \): full
& Try \( T[2] \): full
& Try \( T[4] \): full
& Try \( T[8] \): ok

Search(64) follows same sequence. Would return "not found" after 4 attempts.
ex: \( h(64) \rightarrow 9, 2, 4, 8, 1, 3, 11, 7, 10, 5, 6. \)

Really this is \( h(k, i) \)

\[
\begin{align*}
h(64, 1) &= 9 / h(64, 2) = 2 / h(64, 3) = 4 / \text{etc.}
\end{align*}
\]

Delete(64)

\( h(64, 1) = 9, \) occupied by 2014
\( h(64, 2) = 2, \) occupied by 43
\( h(64, 3) = 4, \) occupied by 78
\( h(64, 4) = 8, \) found 64, DELETE IT.

\[
\text{(problematic)}
\]

Search(103)

\( h(103, 1) = 4, \) occupied by 78
\( h(103, 2) = 8, \) empty: declare 103 not in T.

Could use special "deleted" markers, but search time increases
Typical probing sequences

Linear probing: \( h(k,i) = (h(k,0) + i) \mod m \) \( \sim h(k) \) and wrap around.

... tends to generate clusters.

![Diagram showing linear probing and cluster formation]

probability of extending a cluster

\[
= \frac{|\text{cluster}|}{m}
\]

slows down search
Typical probing sequences

Linear probing: \( h(k, i) = (h(k, 0) + i) \mod m \) \( \sim h(k) \) and wrap around.

... tends to generate clusters.

Quadratic probing: \( h(k, i) = (h(k, 0) + c \cdot i + d \cdot i^2) \mod m \)

Less clustering, need to make sure sequence hits all slots

Both generate \( m \) probe sequences in total

Double hashing: \( h(k, i) = \left( h_1(k) + i \cdot h_2(k) \right) \mod m \)

Each \( k \) has "random" offset

Generates \( O(m^2) \) probe sequences: better

Heuristic: choose \( m = 2^r \) & \( h_2(k) \) is odd.
Analysis of open addressing

Assuming uniform hashing:

- Each key is equally likely to have any of the $m!$ permutations as probe sequence (independent of other keys)

Even though all we have so far is $O(m^2)$

Simple uniform hashing:

For a random $h$, every slot is equally likely
Analysis of Open Addressing

Assuming Uniform Hashing: each key is equally likely to have any of the $m!$ permutations as probe sequence (independent of other keys).

Recall $n < m$, so $\alpha < 1$. Claim: $E[\#\text{probes}] \leq \frac{1}{1-\alpha} \left( \frac{m}{m-n} \right)$ (search)

If true, then for $n \ll m$ $E[\#\text{probes}] = O(1)$

$\therefore n = \frac{1}{2}m \rightarrow 2$ probes

$\therefore 90\%$ full table $\rightarrow 10$ probes

Works well if you can afford a table $\sim$ data $\times 2$
Claim: \( E[\#\text{probes}] \leq \frac{1}{1-\alpha} \)

Look at unsuccessful search

\* \( P[\text{1st probe collides}] = \frac{n}{m} \rightarrow \text{need 2nd probe} \)

\( P[\text{2nd probe collides}] = \frac{n-i}{m-1} \rightarrow \text{need 3rd probe} \)

\[ \vdots \]

\[ \frac{n-i}{m-i} < \frac{n}{m} = \alpha \]

\( E[\#\text{probes}] = 1 + \frac{n}{m} \left( \begin{array}{c} \text{need at least a 2nd probe} \end{array} \right) \)
Claim: $E[\text{#probes}] \leq \frac{1}{1-\alpha}$

Look at unsuccessful search

$P[\text{1st probe collides}] = \frac{n}{m}$ \rightarrow need 2nd probe

$\ast \quad P[\text{2nd probe collides}] = \frac{n-1}{m-1}$ \rightarrow need 3rd probe

$\vdots$

$\frac{n-i}{m-i} < \frac{n}{m} = \alpha$

$E[\text{#probes}] = 1 + \frac{n}{m} \left( 1 + \frac{n-1}{m-1} \right)$

(need a 3rd probe)

Remember, probe sequence is a permutation.
Never check one slot twice.
Claim: \( E[\text{#probes}] \leq \frac{1}{1-\alpha} \)

Look at unsuccessful search

\[
P[1st \text{ probe collides}] = \frac{n}{m} \quad \rightarrow \text{need 2nd probe}
\]

\[
P[2nd \text{ probe collides}] = \frac{n-1}{m-1} \quad \rightarrow \text{need 3rd probe}
\]

\[
\vdots
\]

\[
\frac{n-i}{m-i} \leq \frac{n}{m} = \alpha
\]

\[
E[\text{#probes}] = 1 + \frac{n}{m} \left( 1 + \frac{n-1}{m-1} \left( 1 + \frac{n-2}{m-2} \left( \cdots \left( 1 + \frac{1}{m-\alpha} \right) \right) \right) \right) \\
\leq 1 + \alpha \left( 1 + \alpha \left( 1 + \alpha \left( \cdots \left( 1 + \alpha \right) \right) \right) \right) \quad \cdots \text{n terms}
\]

\[
\leq 1 + \alpha + \alpha^2 + \alpha^3 + \cdots \quad \cdots \infty \text{ terms}
\]

\[
= \sum_{i=0}^{\infty} \alpha^i = \frac{1}{1-\alpha}
\]

Remember, probe sequence is a permutation.

Never check one slot twice.

See CLRS for alternate analysis incl. successful search