CS 225
Data Structures

March 15 – Hash Table Collisions
Wade Fagen-Ulmschneider, Craig Zilles
Collision Handling: Separate Chaining

\( S = \{16, 8, 4, 13, 29, 11, 22\} \) \quad |S| = n

\( h(k) = k \mod 7 \) \quad |Array| = N

<table>
<thead>
<tr>
<th>Worst Case</th>
<th>SUHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert</td>
<td></td>
</tr>
<tr>
<td>Remove/Find</td>
<td></td>
</tr>
</tbody>
</table>
Collision Handling: Probe-based Hashing

$S = \{ 16, 8, 4, 13, 29, 11, 22 \}$ \quad |S| = n

$h(k) = k \% 7$ \quad |Array| = N

(Example of closed hashing)
Collision Handling: Linear Probing

\[ \text{S} = \{ 16, 8, 4, 13, 29, 11, 22 \} \quad |\text{S}| = n \]
\[ h(k) = k \mod 7 \quad |\text{Array}| = N \]

Try \( h(k) = (k + 0) \mod 7 \), if full...
Try \( h(k) = (k + 1) \mod 7 \), if full...
Try \( h(k) = (k + 2) \mod 7 \), if full...
Try ...

(Example of closed hashing)

<table>
<thead>
<tr>
<th></th>
<th>Worst Case</th>
<th>SUHA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insert</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Remove/Find</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
A Problem w/ Linear Probing

Primary clustering:

Description:

Remedy:
Collision Handling: Double hashing

\[ S = \{ 16, 8, 4, 13, 29, 11, 22 \} \quad |S| = n \]
\[ h(k) = k \mod 7 \]
\[ |Array| = N \]

Try \( h(k) = (k + 0 \times h_2(k)) \mod 7 \), if full...
Try \( h(k) = (k + 1 \times h_2(k)) \mod 7 \), if full...
Try \( h(k) = (k + 2 \times h_2(k)) \mod 7 \), if full...
Try ...

\[ h(k, i) = (h_1(k) + i \times h_2(k)) \mod 7 \]
Running Times

The expected number of probes for find(key) under SUHA

Linear Probing:
• Successful: $\frac{1}{2}(1 + \frac{1}{1-\alpha})$
• Unsuccessful: $\frac{1}{2}(1 + \frac{1}{(1-\alpha)^2})$

Double Hashing:
• Successful: $\frac{1}{\alpha} \times \ln(\frac{1}{1-\alpha})$
• Unsuccessful: $\frac{1}{1-\alpha}$

Separate Chaining:
• Successful: $1 + \frac{\alpha}{2}$
• Unsuccessful: $1 + \alpha$

(Don’t memorize these equations, no need.)

Instead, observe:
- As $\alpha$ increases:
- If $\alpha$ is constant:
Running Times

The expected number of probes for find(key) under SUHA

**Linear Probing:**
- Successful: $\frac{1}{2}(1 + \frac{1}{(1-\alpha)})$
- Unsuccessful: $\frac{1}{2}(1 + \frac{1}{(1-\alpha)})^2$

**Double Hashing:**
- Successful: $\frac{1}{\alpha} \times \ln\left(\frac{1}{(1-\alpha)}\right)$
- Unsuccessful: $\frac{1}{(1-\alpha)}$
ReHashing

What if the array fills?
Which collision resolution strategy is better?
- Big Records:

- Structure Speed:

What structure do hash tables replace?

What constraint exists on hashing that doesn’t exist with BSTs?

Why talk about BSTs at all?