

Every hash table contains three pieces:

1. A **hash function, $f(k)$** . The hash function transforms a key from the keyspace into a small integer.
2. **An array**.
3. **A mystery** third element.

Collision Handling Strategy #2: Probe-based Hashing

Example: $S = \{ 16, 8, 4, 13, 29, 11, 22 \}$, $|S| = n$
 $h(k) = k \% 7$, $|Array| = N$

[0]	
[1]	
[2]	
[3]	
[4]	
[5]	
[6]	
[7]	

Linear Probing:

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

Linear Probing leads to Primary Clustering

Description:

Remedy:

Double Hashing:

Example: $S = \{ 16, 8, 4, 13, 29, 11, 22 \}$, $|S| = n$
 $h_1(k) = k \% 7$, $h_2(k) = 5 - (k \% 5)$, $|Array| = N$

[0]	
[1]	
[2]	
[3]	
[4]	
[5]	
[6]	
[7]	

Double Hashing:

Try $h(k) = (k + 0 * h_2(k)) \% 7$, if full...
 Try $h(k) = (k + 1 * h_2(k)) \% 7$, if full...
 Try $h(k) = (k + 2 * h_2(k)) \% 7$, if full...
 ...

$$h(k, i) = (h_1(k) + i * h_2(k)) \% 7$$

Running Time:

Linear Probing:

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

Double Hashing:

- Successful: $\frac{1}{\alpha} * \ln(\frac{1}{(1-\alpha)})$
- Unsuccessful: $\frac{1}{(1-\alpha)}$

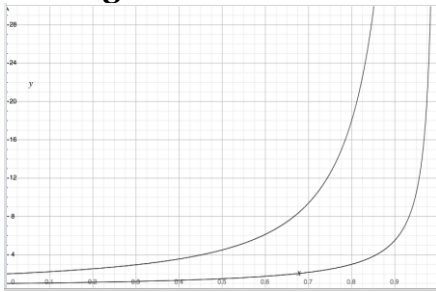
Separate Chaining:

- Successful: $1 + \alpha/2$
- Unsuccessful: $1 + \alpha$

Running Time Observations:

1. As α increases:
2. If α is held constant:

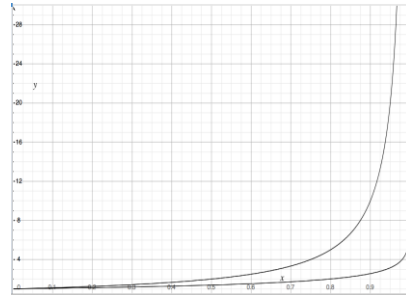
Running Time Observations:



Linear Probing:

Successful: $\frac{1}{2}(1 + 1/(1-\alpha))$

Unsuccessful: $\frac{1}{2}(1 + 1/(1-\alpha))^2$



Double Hashing:

Successful: $1/\alpha * \ln(1/(1-\alpha))$

Unsuccessful: $1/(1-\alpha)$

Analysis of Dictionary-based Data Structures

	Hash Table		AVL	List
	Amortized	Worst Case		
Find				
Insert				
Storage Space				

A Secret, Mystery Data Structure:

ReHashing:

What happens when the array fills?

Better question:

Algorithm:

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?

ADT:

insert

remove

isEmpty

CS 225 – Things To Be Doing:

1. Theory Exam 3 starts next week (Tuesday, April 3rd)
2. MP5 has been released; EC+7 deadline is tonight
3. lab_btree due this Tuesday, March 27th at 11:59pm
4. Daily POTDs are ongoing!