DIT181: Data Structures and Algorithms

Hash Tables

Gül Calikli
Email: calikli@chalmers.se

Hash Tables

- Where to we use Hash Tables?
  - Database indexing,
  - Program compilation (i.e., hash tables keep track of declared variables in source codes),
  - Password authentication,
  - Computer game programs,
  - Online spelling checkers, etc.

Hash Tables

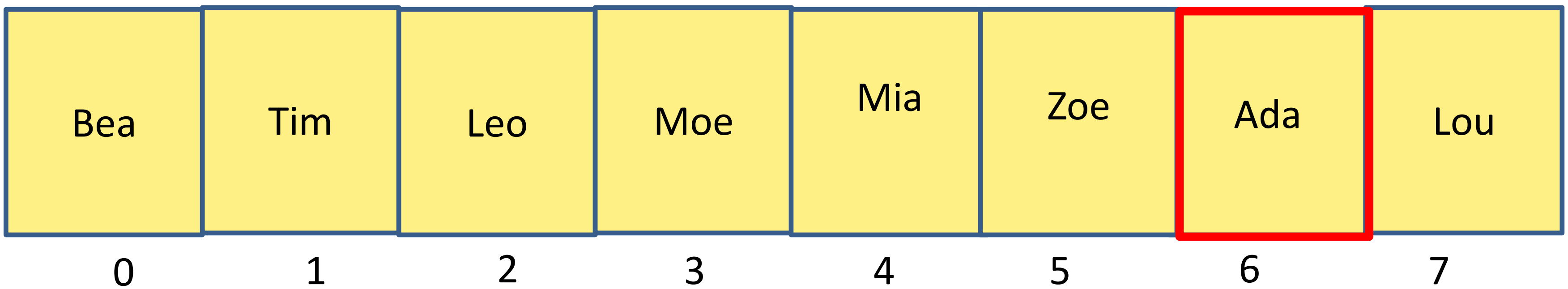
- Consider the following array. How would you find “Ada” in the array?

Bea	Tim	Leo	Moe	Mia	Zoe	Ada	Lou

Answer: Through linear search in array (Time Complexity: O(N)
→ If the array size is big this could take long time!!!)

Hash Tables

- How would you find the item “Ada” in the array, if you know the index number of that item?

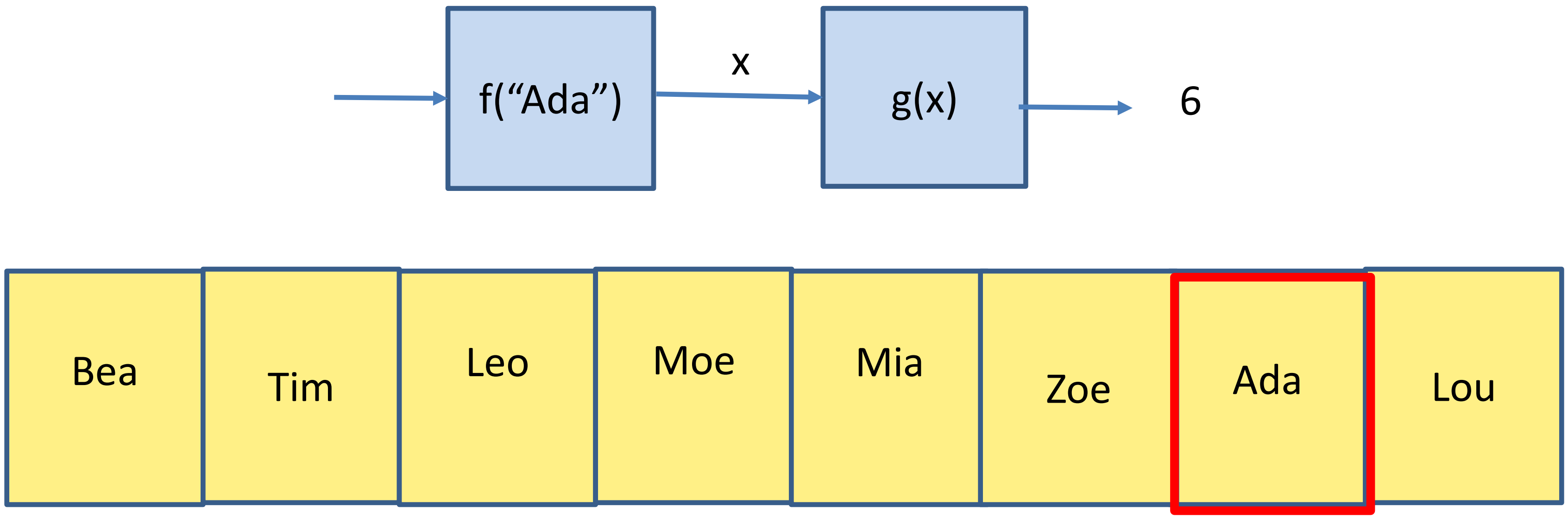


Answer: Find Ada = 6
MyData = Array[6]

Time Complexity: O(1)

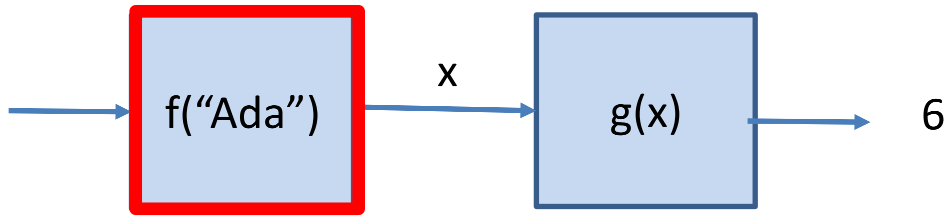
Hash Tables

- But how can you know which element of the array contains the element, you are looking for?
- You can calculate the index using the value of the item itself!



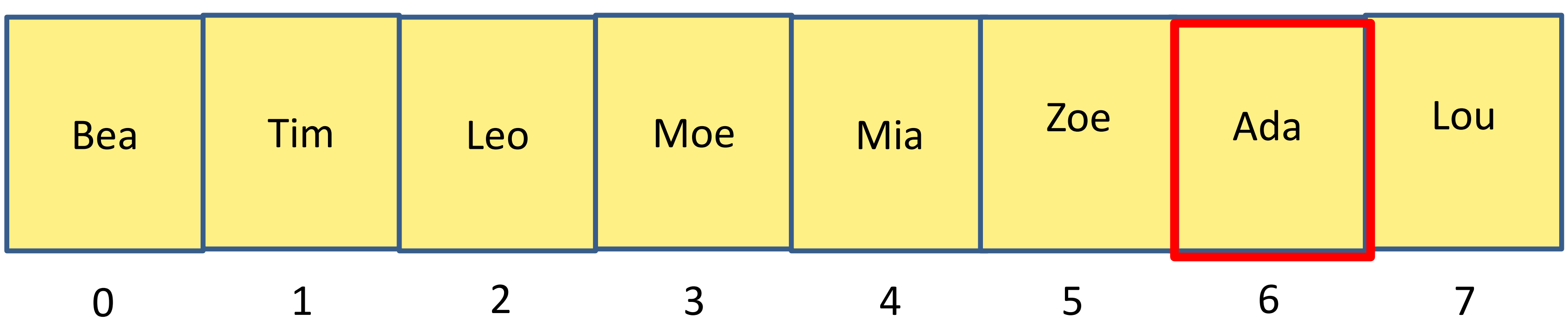
Hash Tables

- But how can you know which element of the array contains the element, you are looking for?
- You can calculate the index using the value of the item itself!



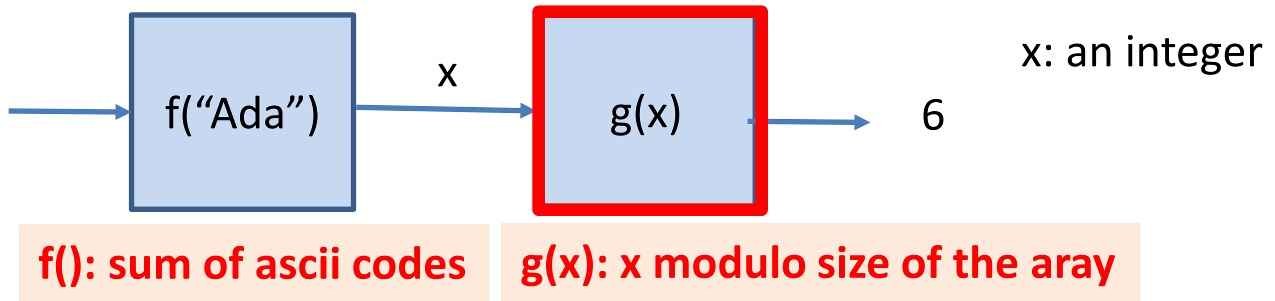
For example, take ASCII code of each character in string "Ada" and sum them up:

A → 65 D → 100 A → 97 f("Ada") = 262



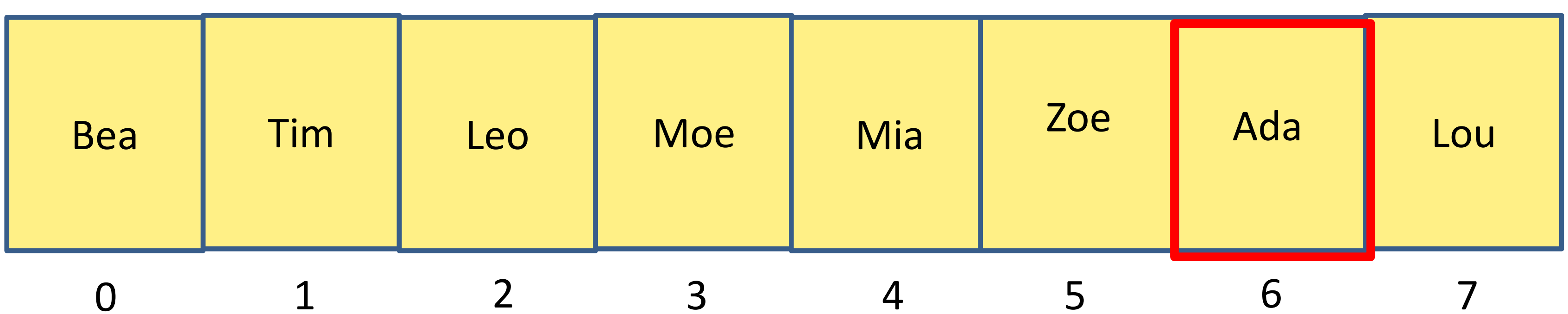
Hash Tables

- But how can you know which element of the array contains the element, you are looking for?
- You can calculate the index using the value of the item itself!



… then divide the result x with the number of elements in the array:

262 % 8 = 6



Hash Tables

- You do not need to store only individual items in a hash table.

Ada
13.03.1981
Computer Scientist
British

In object oriented approach, "**Ada**" might be the key to find the index of the object on the array

Bea 24.07.1978 Architect Dutch	Tim 04.01.1983 Bartender Irish	Leo 19.05.1976 Physicist Polish	Moe 27.04.169 Politician American	Mia 17.08.1973 Baker German	Zoe 16.03.1974 Lawyer British	Ada 15.04.1981 Computer Scientist British	Lou 15.04.1980 Computer Scientist French

Hashing Algorithm

- A hashing algorithm is a calculation applied to a key, which might be a very long string or a very large number to transform it into a relatively a small index number that corresponds to a position in the hash table.
- Some Examples:
- For numeric keys:
  - Address = key mod n
- For alphanumeric keys:
  - Address = (Sum of ascii codes of each character in the key) modulo n

Collisions

- Assume that you want to place object with key “Mia” into the hash table

Mia
17.08.1973 Baker German

Add all ascii codes in “Mia”: 83+ 117+ 101 = 301

Take modulo 8 → 301 % 8 = 5

Bea	Tim	Leo	Moe	Zoe		Ada	
24.07.1978 Architect Dutch	04.01.1983 Bartender Irish	19.05.1976 Physicist Polish	27.04.169 Politician American	16.03.1974 Lawyer British		15.04.1981 Computer Scientist British	

Collisions

- Assume that you want to place object with key “Mia” into the hash table

Mia 17.08.1973 Baker German

Add all ascii codes in “Mia”: 83+ 117+ 101 = 301

Take modulo 8 → 301 % 8 = 5

Bea 24.07.1978 Architect Dutch	Tim 04.01.1983 Bartender Irish	Leo 19.05.1976 Physicist Polish	Moe 27.04.169 Politician American	Zoe 16.03.1974 Lawyer British	Mia 17.08.1973 Baker German	Ada 15.04.1981 Computer Scientist British	

Collisions

- Now you want to insert “Sue” into the table

Add all ascii codes in “Sue”: 77+ 105+ 97 = 264
Take modulo 8→ 264 % 8 = 0

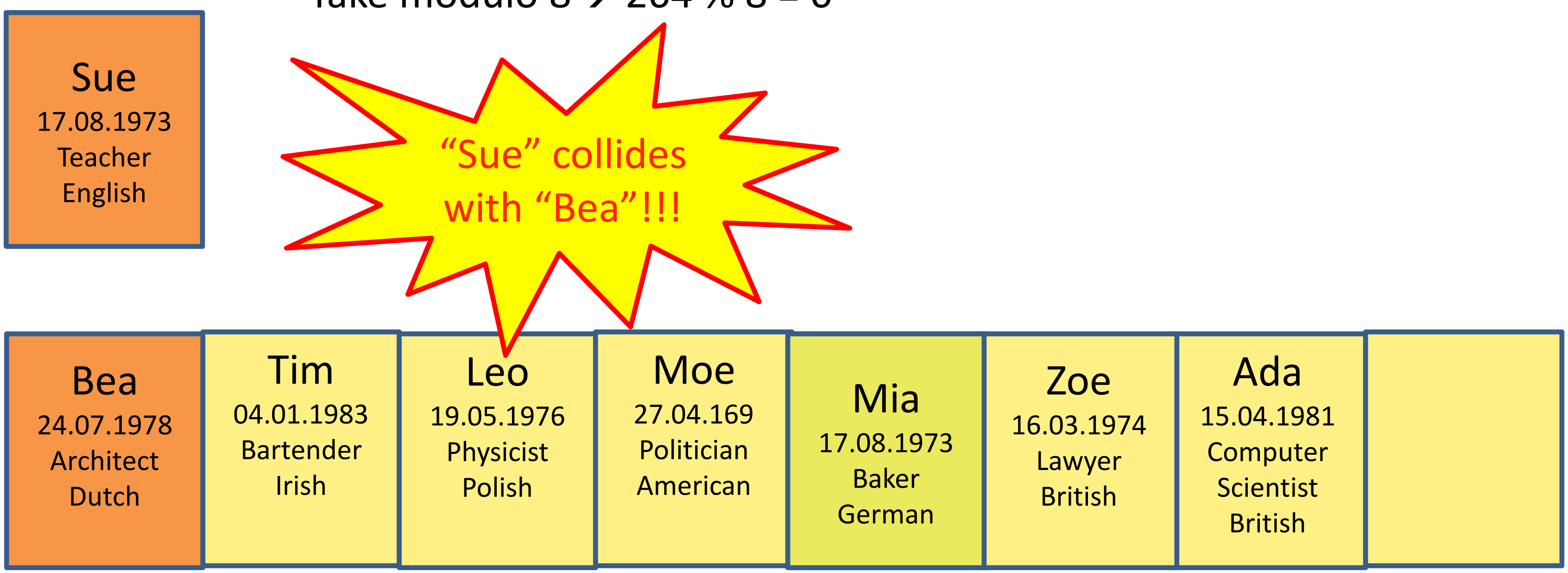
Sue 17.08.1973 Teacher English

Bea 24.07.1978 Architect Dutch	Tim 04.01.1983 Bartender Irish	Leo 19.05.1976 Physicist Polish	Moe 27.04.169 Politician American	Mia 17.08.1973 Baker German	Zoe 16.03.1974 Lawyer British	Ada 15.04.1981 Computer Scientist British	

Collisions

- Now you want to insert “Sue” into the table

Add all ascii codes in “Sue”: 77+ 105+ 97 = 264
Take modulo 8→ 264 % 8 = 0



Linear Probing

- Now you want to insert “Sue” into the table

Sue
17.08.1973
Teacher
English

Linear Probing: Search for the next available spot in the array and place the object with key “Sue” there

Bea 24.07.1978 Architect Dutch	Tim 04.01.1983 Bartender Irish	Leo 19.05.1976 Physicist Polish	Moe 27.04.169 Politician American	Mia 17.08.1973 Baker German	Zoe 16.03.1974 Lawyer British	Ada 15.04.1981 Computer Scientist British	

Linear Probing

- Now you want to insert “Sue” into the table

Sue
17.08.1973
Teacher
English

Linear Probing: Search for the next available spot in the array and place the object with key “Sue” there

Bea 24.07.1978 Architect Dutch	Tim 04.01.1983 Bartender Irish	Leo 19.05.1976 Physicist Polish	Moe 27.04.169 Politician American	Mia 17.08.1973 Baker German	Zoe 16.03.1974 Lawyer British	Ada 15.04.1981 Computer Scientist British	Sue 17.08.1973 Teacher English

Lazy Deletion

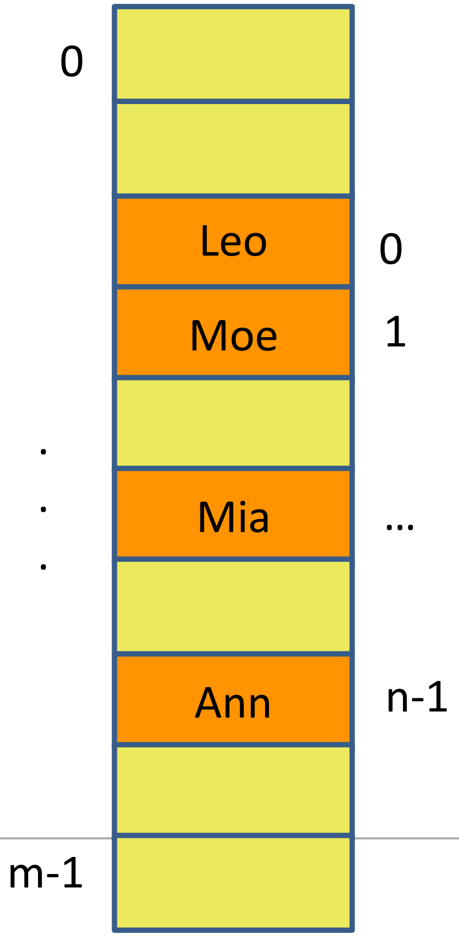
- Emptying slots can break probe sequence and could cause **`find()`** operation stop prematurely during collision resolution
- An item in a hash table not only represents itself but also it also connects other elements by serving as a placeholder during collision resolution
- Hence, we do lazy deletion, by marking elements as “deleted” rather than physically removing them.

Load Factor

- The load factor λ of a probing hash table is the fraction of the table that is full.
- The load factor λ ranges from 0 (empty) to 1 (completely full).

$$\lambda = n/m$$

- n: number of elements stored in the table
- m: number of slots in the table



Primary Clustering

- As long as table table is big enough, a free cell can always be found, but time to do so can get quite long.
- Worse, even of the table is relatively empty, blocks of occupied cells start forming.
- This effect is known as primary clustering.

Quadratic Probing

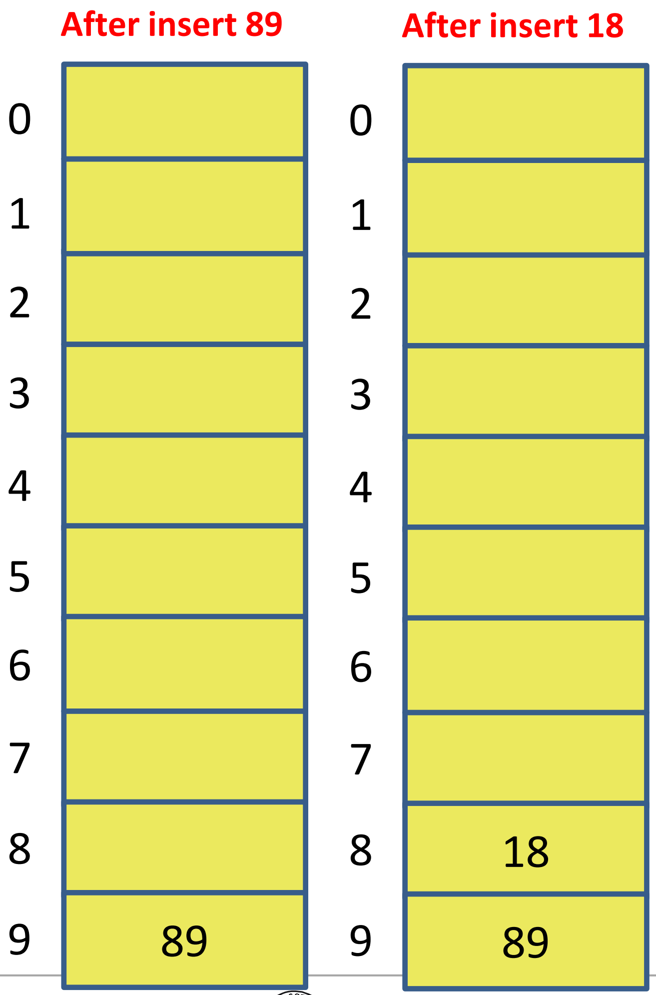
- If the hash function evaluates to H, and a search in cell H is inconclusive, we try cells $H + i^2$, where $i = 1, 2, 3, 4, \ldots$.

hash(89, 0) = 9

0	
1	
2	
3	
4	
5	
6	
7	
8	
9	89

Quadratic Probing

- If the hash function evaluates to H, and a search in cell H is inconclusive, we try cells $H + i^2$, where $i = 1, 2, 3, 4, \ldots$.



hash(89, 0) = 9
hash(18,10) = 8

Quadratic Probing

- If the hash function evaluates to H, and a search in cell H is inconclusive, we try cells $H + i^2$, where $i = 1, 2, 3, 4, \ldots$.

	After insert 89	After insert 18	After insert 49 & 58
0			49
1			
2			58
3			
4			
5			
6			
7			
8		18	18
9	89	89	89

hash(89,10) = 9
hash(18,10) = 8
hash(49,10) = 9
hash(58,10) = 8

Quadratic Probing

- If the hash function evaluates to H, and a search in cell H is inconclusive, we try cells $H + i^2$, where $i = 1, 2, 3, 4, \ldots$.

	After insert 89	After insert 18	After insert 49 & 58	After insert 9
0			49	49
1				
2			58	58
3				9
4				
5				
6				
7				
8		18	18	18
9	89	89	89	89

hash(89, 0) = 9
hash(18,10) = 8
hash(49,10) = 9
hash(58,10) = 8
hash(9, 10) = 9

Quadratic Probing

- Theorem: If **quadratic probing** is used and table size is a prime number, then a new element can always be inserted if the table is at least half empty ($\lambda \leq 0.5$). Furthermore, no cell is probed twice.

Quadratic Probing

- Theorem: If **quadratic probing** is used and table size is a prime number, then a new element can always be inserted if the table is at least half empty ($\lambda \leq 0.5$). Furthermore, no cell is probed twice.

Proof Strategy:

- Let **M** be the **size of the table**
- Assume that **M** is a **prime number** and **M > 3**
- We will show that the first $\lceil M/2 \rceil$alternative locations are distinct

Quadratic Probing

- Theorem: If **quadratic probing** is used and table size is a prime number, then a new element can always be inserted if the table is at least half empty ($\lambda \leq 0.5$). Furthermore, no cell is probed twice.

Proof Strategy:

- Let **M** be the **size of the table**
- Assume that **M** is a **prime number** and **M > 3**
- We will show that the first $\lceil M/2 \rceil$alternative locations are distinct

So Let's show the following:

- We will show that the first $\lceil M/2 \rceil$alternative locations are distinct

Quadratic Probing

- Theorem: If **quadratic probing** is used and table size is a prime number, then a new element can always be inserted if the table is at least half empty ($\lambda \leq 0.5$). Furthermore, no cell is probed twice.

Proof Strategy:

- Let **M** be the **size of the table**
- Assume that **M** is a **prime number** and **M > 3**
- We will show that the first $\lceil M/2 \rceil$alternative locations are distinct

Let $H + i^2 \ (mod\ M)$ and $H + j^2 \ (mod\ M)$, be two identical locations, where $0 \leq i, j \leq \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$.

Quadratic Probing

GOAL: We will show that the first $[M/2]$ alternative locations are **distinct.**

Let $H + i^2 \ (mod\ M)$ and $H + j^2 \ (mod\ M)$, be two identical locations, where $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$.

Then:

$H + \ i^2 \equiv \ H + \ j^2$ (mod M)

$i^2 \equiv \ j^2$ (mod M)

$i^2 - j^2$) $\equiv$ 0 (mod M)

$(i - j)(i + j) \equiv$ 0 (mod M)

Quadratic Probing

GOAL: We will show that the first $\lceil M/2 \rceil$ alternative locations are **distinct.**

Let $H + i^2 \ (mod\ M)$ and $H + j^2 \ (mod\ M)$, be two identical locations, where $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and $i \ne j$.

Then:

$H + i^2 \equiv H + j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

$i^2 - j^2 \equiv 0$(mod M)

$\boxed{(i - j)(i + j) \equiv 0 \text{ (mod M)}}$

This follows that either i-j or i+j is divisible by M

Quadratic Probing

GOAL: We will show that the first $\lceil M/2 \rceil$ alternative locations are **distinct.**

Let $H + i^2\ (mod\ M)$ and $H + j^2\ (mod\ M)$, be two identical locations, where $0 \leq i, j \leq \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$.

Then:

$H + i^2 \equiv H + j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

$i^2 - j^2 \equiv 0$ (mod M)

$\boxed{(i - j)(i + j) \equiv 0 \text{ (mod M)}}$

$i \neq j$, hence $i - j \neq 0$

Quadratic Probing

GOAL: We will show that the first $\lceil M/2 \rceil$ alternative locations are **distinct.**

Let $H + i^2\ (mod\ M)$ and $H + j^2\ (mod\ M)$, be two identical locations, where $0 \leq i, j \leq \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$.

Then:

$H + i^2 \equiv H + j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

$i^2 - j^2 \equiv 0$ (mod M)

$\boxed{(i - j)(i + j) \equiv 0 \text{ (mod M)}}$

Also, $\boldsymbol{i + j < M}$, since $0 \leq i, j \leq \left\lfloor \frac{M}{2} \right\rfloor$ and M is a prime number.

Hence, $\boldsymbol{i + j < M}$ and $\boldsymbol{i - j < M}$

As a result, $\boldsymbol{i + j}$ or $\boldsymbol{i - j}$ **cannot** be divisible by M.

Quadratic Probing

GOAL: We will show that the first $[M/2]$ alternative locations are **distinct.**

Let $H + i^2\ (mod\ M)$ and $H + j^2\ (mod\ M)$, be two identical locations, where $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$.

Then:

$H + \ i^2 \ = \ H + \ j^2$ (mod M)

$i^2 \ = \ j^2$ (mod M)

$i^2 - j^2$ (mod M)

$\boxed{(i - j)(i + j) = 0 \text{ (mod M)}}$

This is a contradiction!!

Also, $\boldsymbol{i + j} \ < \boldsymbol{M}$, since $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and M is a prime number.

Hence, $\boldsymbol{i + j}$ or $\boldsymbol{i - \ j}$ **cannot** be divisible by M.

Quadratic Probing

GOAL: We will show that the first $\lceil M/2 \rceil$ alternativ… ocations are **distinct.**

Let $H + i^2\ (mod\ M)$ an… …cal

locations, where 0…

Then:

$H + i^2 \equiv H + j^2$ (m…

This is a contradiction!!
Hence the following NEVER holds
for $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and $i \ne j$;
$H + i^2 \equiv H + j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

$\boxed{(i - j)(i + j) \equiv 0 \text{ (mod M)}}$

Also, …$+ j < M$, since $0 \le i, j \le \left\lfloor \frac{M}{2} \right\rfloor$ and M is a prime number.

Hence, $i + j$ or $i - j$ **cannot** be divisible by M

Quadratic Probing

GOAL: We will show t… the first $\lceil M/$… alternati… cations are **distinct.**

Let $H + i^2\ (mod\ M$…

locations, where $0 \leq i, j$…

Then:

$H + i^2 \equiv H + j^2$ (mod M)

$i^2 \equiv j^2$ (mod M)

This is a contradiction!!
Hence the following NEVER holds
for $0 \leq i, j \leq \left\lfloor \frac{M}{2} \right\rfloor$ and $i \neq j$;
$H + i^2 \equiv H + j^2$ (mod M)

Also …

…M is a prime

***j* cannot** be

divisible by M

- **THUS, the following theorem has been proven**
- Theorem: If **quadratic probing** is used and table size is a prime number, then a new element can always be inserted if the table is at least half empty ($\lambda \leq 0.5$). Furthermore, no cell is probed twice.