

# Constant Expected Chain Lengths



# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.

# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.  
 $\Rightarrow$  The expected length of a chain is  $\frac{n}{m}$ .



# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.  
 $\implies$  The expected length of a chain is  $\frac{n}{m}$ .
- Because  $m \approx n$  the expected chain length is constant.

# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.  
 $\implies$  The expected length of a chain is  $\frac{n}{m}$ .
- Because  $m \approx n$  the expected chain length is constant.
- However, this argument only applies if:



# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.
  - $\Rightarrow$  The expected length of a chain is  $\frac{n}{m}$ .
- Because  $m \approx n$  the expected chain length is constant.
- However, this argument only applies if:
  - The hash function is well-chosen.

# Constant Expected Chain Lengths



- Recall we expect to see only  $n \ll |\mathcal{U}|$  of the keys.
  - $\implies$  The expected length of a chain is  $\frac{n}{m}$ .
- Because  $m \approx n$  the expected chain length is constant.
- However, this argument only applies if:
  - The hash function is well-chosen.
  - The data encountered is non-pathological.





# The Existence of Long Chains





# The Existence of Long Chains

- However, the *worst-case* chain length can be huge, i.e.  $\Omega(n)$  if:



# The Existence of Long Chains

- However, the *worst-case* chain length can be huge, i.e.  $\Omega(n)$  if:
  - The hash function is not well-chosen.





# The Existence of Long Chains

- However, the *worst-case* chain length can be huge, i.e.  $\Omega(n)$  if:
  - The hash function is not well-chosen.
  - Or, the *data encountered* has patterns or structural flaws.





## The Existence of Long Chains

- However, the *worst-case* chain length can be huge, i.e.  $\Omega(n)$  if:
  - The hash function is not well-chosen.
  - Or, the *data encountered* has patterns or structural flaws.
- Indeed, **Denial of Service** attacks often exploit such flaws.

# Universal Hash Functions



# Universal Hash Functions

- So for hashing to work well we must keep the chains short.



# Universal Hash Functions

- So for hashing to work well we must keep the chains short.
  - That is, we must try to *avoid* collisions.

# Universal Hash Functions

- So for hashing to work well we must keep the chains short.
  - That is, we must try to *avoid* collisions.
- The best way to do this is via a **universal** hash function.



# Universal Hash Functions

- So for hashing to work well we must keep the chains short.
  - That is, we must try to *avoid* collisions.
- The best way to do this is via a **universal** hash function.
- A function  $h$  has the universal hash function property if:

$$\mathbb{P} \left( h(k) = h(\hat{k}) \right) = \frac{1}{m} \quad \forall k \neq \hat{k}$$



# Universal Hash Functions

- So for hashing to work well we must keep the chains short.
  - That is, we must try to *avoid* collisions.
- The best way to do this is via a **universal** hash function.
- A function  $h$  has the universal hash function property if:

$$\mathbb{P} \left( h(k) = h(\hat{k}) \right) = \frac{1}{m} \quad \forall k \neq \hat{k}$$

- This is the best we can hope for, but do universal hash functions exist?