

# Collision Resolution

CS240

# Reminder: Hash Functions

- Hash function maps a key to a valid array index.
  - $h(k) \rightarrow [0, N - 1]$
- Accomplished in two steps:
  - **Hash code** converts the key to a 32 or 64 bit integer
    - For short, fixed-sized keys: integer representation of bits
    - Otherwise: Polynomial hash code, Cyclic bit-shifting, etc.
  - **Compression function** maps the integer to an array index
    - Division method:  $i \% N$
    - Multiply-and-divide:  $((ai + b) \% p) \% N$

# Collision Resolution: Separate Chaining

- Each bucket contains a linked list of (key, value) pairs
- The **load factor**  $\lambda = n/N$  describes how full the table is
- Average lookup time is  $O(\lambda)$
- Therefore we should make sure that  $\lambda \in O(1)$

# Collision Resolution: Open Addressing

- All item references are stored directly in available table slot
- On collision, a search is conducted for a free slot
- Linear probing:  $A[(h(k) + i) \% N]$
- Quadratic probing:  $A[(h(k) + i^2) \% N]$
- Double hashing:  $A[(h(k) + i \times h'(k)) \% N]$

# Behavior of Linear Probing

