Prepared by: Ashik Emon

Problem Statement

Let $U = \mathbb{N}$ be the universe of keys, $T = \{0, 1, \ldots, m-1\}$ the hash table with $m > 1$, and $S \subseteq U$ a randomly selected set of n keys. Assume a hash function $h : U \to T$ satisfies:

$$h(x) = x \bmod m.$$

Collisions are resolved by quadratic probing: if $h(x)$ is occupied, probe $h(x)+i^2$ for $i = 1, 2, \ldots$ until a free slot is found.

Part (a): Is this function well-balanced?

A hash function is *well-balanced* if it distributes keys uniformly across the table and minimizes clustering. The probing sequence here is:

$$h(x),\ h(x)+1^2,\ h(x)+2^2,\ h(x)+3^2, \ldots \pmod m.$$

This is quadratic probing. While it reduces primary clustering compared to linear probing, it does not guarantee visiting all positions unless m is prime and certain conditions hold. Keys with the same initial hash value follow identical probe sequences, causing *secondary clustering*. Furthermore, the distribution depends on m and quadratic increments, which can leave some slots unreachable.

Final Thought: The function is *not well-balanced* because it does not ensure uniform distribution and can lead to clustering.

Part (b): Expected Collisions under Simple Uniform Hashing

Assume simple uniform hashing: for any key $a \in U$ and slot $i \in T$,

$$\Pr(h(a) = i) = \frac{1}{m},$$

with keys hashed independently and uniformly.

We show:

1. The expected number of collisions at some hash value is less than $\frac{n^2}{2m^2}$.
2. If $n = m$, the probability that more than one element from S maps to the same hash value is less than $\frac{1}{2}$.

Defining Indicator Variables

For $i, j \in \{1, \ldots, n\}$ with $i < j$, define:

$$X_{ij} = \begin{cases} 1 & \text{if keys } i \text{ and } j \text{ collide (same hash value)} \\ 0 & \text{otherwise.} \end{cases}$$

The total number of collisions is:

$$X = \sum_{i<j} X_{ij}.$$

Computing Expected Value

Since hashing is uniform and independent:

$$\mathbb{E}[X_{ij}] = \Pr(h(i) = h(j)) = \frac{1}{m}.$$

There are $\binom{n}{2} = \dfrac{n(n-1)}{2}$ pairs, so:

$$\mathbb{E}[X] = \sum_{i<j} \mathbb{E}[X_{ij}] = \binom{n}{2} \cdot \frac{1}{m} = \frac{n(n-1)}{2m} < \frac{n^2}{2m}.$$

Expected Collisions at One Hash Value

For a single slot $t \in T$, define:

$$Y_t = \text{number of keys hashed to } t.$$

Then:

$$\mathbb{E}[Y_t] = \frac{n}{m}.$$

Expected collisions at t:

$$\mathbb{E}\left[\binom{Y_t}{2}\right] \approx \frac{(n/m)^2}{2} = \frac{n^2}{2m^2}.$$

Probability of Collision When $n = m$

Using the Birthday Paradox approximation:

$$\Pr(\text{at least one collision}) \approx 1 - e^{-n(n-1)/(2m)}.$$

For $n = m$:

$$\Pr(\text{collision}) \approx 1 - e^{-1/2} \approx 0.39 < \frac{1}{2}.$$

Final Thought

- Quadratic probing is not well-balanced due to clustering and incomplete coverage.
- Expected number of collisions in the table: $\mathbb{E}[X] < \frac{n^2}{2m}$.
- Expected collisions at one hash value: $< \frac{n^2}{2m^2}$.
- If $n = m$, probability of collision $< \frac{1}{2}$.