Hashing contd.

$h_{a,b}(x): x \to (a x + b) \mod N \mod m$

$a, b, x \in U \quad N \text{ is a prime number}$

$|U| = N$

$m: \text{size of table}$

Chaining method
Perfect hash function

\[ S \rightarrow T \]

We must ensure that no more than one element is mapped to any location.

\[ |T| > |S| \]

\[ \delta_h(x, y) = 1 \text{ if } h(x) = h(y) \]

\[ 0 \text{ otherwise} \]
Using universal hash function, the probability that $h(x) = h(y)$ for a randomly chosen $h \in H$ is
$$\frac{c}{m}$$

If $X$ is a 0, 1 random variable and $\text{prob} (x = 1) = p$
then $E[X] = p$

The total expected # collisions in a set $S$, where $|S| = n$

$$E \left[ \sum_{h \in H} \sum_{x \neq y \in S} \delta_h(x, y) \right] \leq E \left[ \sum_{h \in H} \delta_h(x, y) \right]$$
and hence

$$f = \binom{n}{2} \frac{c}{m}.$$ Then by Markov's inequality $\text{Prob} \cdot \text{that \ the \ no. \ of \ collisions exceed } 2 \cdot f \leq \frac{1}{2}$
For \( C = 2 \), \( m \geq 4n^2 \), the value of \( 2c \) is less than \( \frac{1}{2} \), i.e., no collisions.

With prob \( \frac{1}{2} \), there are no collisions if table size is about \( 2^c n^2 \).

We use a two level strategy:

1. First we hash the elements using a function \( H \).

   (There could be collisions, suppose there are \( n_i \) elements hashed to location \( i \), \( n_i > 0 \).

2. Next level, for elements in location \( i \), use the previous observation, i.e., hash these elements, \( s_i \), using \( 4n_i \) locations.
\[ E \left[ \sum_{i} \eta_i^2 \right] = O(m) \]

see for details in notes