# 8.5 Universal Classes of Hash Functions

**Definition:** Let $K$ be the set of all possible keys, and $H$ a set of functions $\{h \mid h: K \to \{0, 1, \ldots, m-1\}\}$. $H$ is universal iff, for any $x_1 \neq x_2 \in K$,

$$\frac{\#\{h \in H \mid h(x_1) = h(x_2)\}}{\#H} \leq \frac{1}{m}.$$

That is, picking $h \in H$ at random means the probability that $h$ causes $x_1$ and $x_2$ to collide is $\frac{1}{m}$, as good as if the bucket $h(x_2)$ is chosen at random independently of $x_1$.

Review of 2 number-theoretic facts from CSE321:

**Fact 1:** $r_1 \equiv r_2 \pmod m$ iff $m \mid (r_1 - r_2)$
iff $r_1 \bmod m = r_2 \bmod m$

**Fact 2:** If $N$ is prime and $a$ is not a multiple of $N$, then $a$ is invertible mod $N$, i.e.,
$(\exists b)\ ab \equiv 1 \pmod N$.
We use the notation $a^{-1}$ to denote this element $b$.

**Theorem:** Let $K \subseteq \{0, 1, \ldots, N-1\}$, where $N$ is prime. For $a \in \{1, 2, \ldots, N-1\}$ and $b \in \{0, 1, 2, \ldots, N-1\}$, let $h_{a,b}(x) = ((ax+b) \bmod N) \bmod m$.
Then $H = \{h_{a,b} \mid 1 \leq a < N \text{ and } 0 \leq b < N\}$ is a universal class of hash functions.

**Proof:** Fix any $x_1 \neq x_2 \in K$. Let
$A = \{(a, b) \mid 1 \leq a < N \text{ and } 0 \leq b < N\}$ and
$R = \{(r_1, r_2) \mid 0 \leq r_1, r_2 < N \text{ and } r_1 \neq r_2\}$

Let $\varphi(a,b) = (r_1, r_2)$, where
$r_1 = (ax_1 + b) \bmod N$ and $r_2 = (ax_2 + b) \bmod N$.
We will show:
(1) $\varphi : A \to R$.
(2) $\varphi$ is bijective.
(3) $\#\{(r_1, r_2) \in R \mid r_1 \equiv r_2 \pmod m\} \leq \frac{\#H}{m}$.
The theorem will then follow because
$\#\{h_{a,b} \in H \mid h_{a,b}(x_1) = h_{a,b}(x_2)\}$
$= \#\{(a,b) \in A \mid (ax_1 + b) \bmod N \equiv (ax_2 + b) \bmod N \pmod m\}$
$= \#\{(a,b) \in A \mid \varphi(a,b) = (r_1, r_2) \text{ and } r_1 \equiv r_2 \pmod m\}$
$= \#\{(r_1, r_2) \in R \mid r_1 \equiv r_2 \pmod m\} \leq \#H/m$.
(1) All we must show is that $r_1 \neq r_2$. Suppose $r_1 = r_2$.
$r_1 = r_2 \Rightarrow ax_1 + b \equiv ax_2 + b \pmod N$
$\Rightarrow ax_1 \equiv ax_2 \pmod N$
$\Rightarrow x_1 \equiv x_2 \pmod N$ (Fact 2, since $1 \leq a < N$)
$\Rightarrow x_1 \bmod N = x_2 \bmod N$
$\Rightarrow x_1 = x_2$ (since $x_1, x_2 \in K \subseteq \{0, 1, \ldots, N-1\}$)
(2) $|A| = |R| = N(N-1)$, so it suffices to prove $\varphi$ surjective.
That is, $ax_1 + b \equiv r_1 \pmod N$
and $ax_2 + b \equiv r_2 \pmod N$
can be solved for $a \neq 0, b$:
$a(x_1 - x_2) \equiv r_1 - r_2 \pmod N$
$a \equiv (x_1 - x_2)^{-1}(r_1 - r_2) \pmod N \neq 0$ (Fact 2)
$b \equiv r_1 - ax_1 \pmod N$
(3) For each $r_1$, at most $\lceil \frac{N}{m} \rceil - 1 \leq \frac{N+m-1}{m} - 1 = \frac{N-1}{m}$ values of $r_2 \neq r_1$ satisfy $r_1 \equiv r_2 \pmod m$. Thus there are at most $\frac{N(N-1)}{m} = \frac{\#H}{m}$ pairs $(r_1, r_2) \in R$ that satisfy $r_1 \equiv r_2 \pmod m$.