Note on a 2-universal hash family[1]

Assuming every key k that could get hashed belongs to $[0, p-1]$,
$\mathcal{H}_{pm} = \{h_{ab}(k) = ((ak+b) \mod p) \mod m : a \in \{1, 2, \ldots, p-1\}, b \in \{0, 1, \ldots, p-1\}$ for a prime $p > m\}$ is an universal hash family.

Proof:

Let k, l be two distinct keys. Also, let $r = (ak+b) \mod p$ and $s = (al+b) \mod p$. —— (0)

- we know, $r - s \equiv a(k-l) \mod p$

  since $a \neq 0$ (from the definition of $\mathcal{H}_{pm}$), and since $k - l \not\equiv 0 \mod p$ (as every k that hashed $\in [0, p-1]$), $r - s \not\equiv 0 \mod p$

  implying $r \neq s$ —— (1)

  * given (1), the maximum number of distinct (r, s) pairs is $p(p-1)$
  * also, given (1), after applying $\mod p$ in h_{ab} (named level 1 hashing), there is no collision between r and s

- given (k, l) and (a, b), there is a unique (r, s)

  * by rewriting (0), i.e., by expressing a and b in terms of (k, l) and (r, s)

    $(a = (r-s)((k-l)^{-1} \mod p) \mod p$ and $b = (r - ak) \mod p)$,

    we know that there is a unique (a, b) that corresponds to $((k, l), (r, s))$
  * hence, each of the possible $p(p-1)$ choices for the pair (a, b) with $a \neq 0$ yields a distinct (r, s) pair with $r \neq s$
  * in othere words, (r, s) tuples and (a, b) tuples have the correspondance and the number of (r, s) pairs that result is $p(p-1)$ —— (2)
  * for a level 1 hash function f, it is $f((k, l), (a, b)) = (r, s)$; for a fixed (k, l) and (a, b) chosen uniformly at random from $p(p-1)$ possible values is equivalent to (r, s) being chosen untiformly at random from $p(p-1)$ possible values

- from (1) and (2), the probability that the distinct keys k and l may cause a collision is equal to the probability that $r \equiv s \mod m$ —— (3)

  * for a fixed r, due to (1), there are $p-1$ possible values for s —— (4)
  * for a fixed r, there are at most $\lceil \frac{p}{m} \rceil$ values that are congruent to $r \mod m$
  * given that $r \neq s$, s and r together cause a collision whenever s takes any value other than r among these values; hence, s can assume any value among at most $\lceil \frac{p}{m} \rceil - 1 \leq \frac{p-1}{m}$ values —— (5)
  * from (4) and (5), and since r could take any of the p values, $pr(r \equiv s \mod m)$ is at most $\frac{\frac{p(p-1)}{m}}{p(p-1)} = \frac{1}{m}$
  * from (3), $pr(h(k) = h(l)) = pr(r \equiv s \mod m) \leq \frac{1}{m}$

[1] note by R. Inkulu, http://www.iitg.ac.in/rinkulu/