UNIVERSITÀ DI PISA

First hands-on: Universal Hash Family

Algorithm Design (2021/2022)

Gabriele Pappalardo
Email: g.pappalardo4@studenti.unipi.it
Department of Computer Science

February 2022

1 Introduction

1.1 Terminologies and Definitions

This subsection defines terminology and definitions in order to fully understand the following essay.

With the symbol U we denote the set of possible keys. The sets Z_p^*, Z_p are defined as $Z_p^* = \{0, \dots, p-1\}, Z_p = \{1, \dots, p-1\}$ where p is a *prime* number.

Definition 1.1 (Universal Hash Family)**.** *A hash family $\mathcal{H}$ is said to be **universal** if given two different keys k_1 and k_2 the probability of collision is less than $\frac{1}{m}$.*

$$\forall k_1, k_2 \in U, k_1 \neq k_2. Pr_{h \in \mathcal{H}}(\{h(k_1) = h(k_2)\}) \leq \frac{1}{m}$$

1.2 Problem definition

The first Algorithm Design hands-on requests to prove that given the following family of functions:

$$\mathcal{H} := \{h_{ab}(x) = ((ax + b) \bmod p) \bmod m \mid a \in Z_p^*, b \in Z_p\}$$

is *universal* with $m > 1$, $p \in [m+1, 2m)$ *prime*. That is, for any $k_1 \neq k_2$, it holds that:

$$|\mathcal{B}| = |\{h \in \mathcal{H} \mid h(k_1) = h(k_2)\}| = \frac{|\mathcal{H}|}{m}$$

Proof. Two different keys k_1, k_2, collides if and only if the hash function image is equal. So given a function $h_{ab} \in \mathcal{H}$, and $k_1 \neq k_2 \in Z_p$:

$$h_{ab}(k_1) = h_{ab}(k_2) \iff ((ak_1 + b) \bmod p) \bmod m = ((ak_2 + b) \bmod p) \bmod m$$

Consider first r and s, defined as:

$$r = (ak_1 + b) \bmod p$$
$$s = (ak_2 + b) \bmod p$$

If the keys k_1 and k_2 are different, we can see that $r \neq s$. To show this, subtract r from s:

$$r - s \equiv (ak_1 + b) - (ak_2 + b) \pmod p \iff r - s \equiv a(k_1 - k_2) \pmod p$$

In order to be equal, the values r and s should be congruent to 0 in mod$p$. We know from our hypothesis that $k_1, k_2 \in Z_p \wedge k_1 \neq k_2$, therefore $k_1 - k_2 \not\equiv 0 \pmod{p}$. Since by definition $a \in Z_p$ we can conclude that $a \not\equiv 0 \pmod{p}$, thus $r - s \not\equiv 0 \pmod{p}$.

Knowing that value r is different from s, we can derive that a and b are uniquely determined. As a matter of fact, we know:

$$\begin{cases} r \equiv ak_1 + b & (\text{mod } p) \\ s \equiv ak_2 + b & (\text{mod } p) \end{cases} \iff \begin{cases} b \equiv r - ak_1 & (\text{mod } p) \\ s \equiv ak_2 + (r - ak_1) & (\text{mod } p) \end{cases}$$

$$\begin{cases} b \equiv r - ak_1 & (\text{mod } p) \\ s - r \equiv ak_2 - ak_1 & (\text{mod } p) \end{cases} \iff \begin{cases} b \equiv r - ak_1 & (\text{mod } p) \\ s - r \equiv a(k_2 - k_1) & (\text{mod } p) \end{cases}$$

$$\iff \begin{cases} b \equiv r - ak_1 & (\text{mod } p) \\ a \equiv (k_2 - k_1)^{-1}(r - s) & (\text{mod } p) \end{cases}$$

Thus, there is one-to-one correspondence between the pairs (a, b), with $a \neq 0$, and the pair (r, s), with $r \neq s$. We have $p(p-1)$ possibilities to select the pair (a, b) and (r, s).

Therefore, the probability that keys k_1 and k_2 collide is equal to the probability that $r \equiv s \pmod{m}$. For a fixed r, the number to choose s, with $s \neq r, s \equiv r \pmod{m}$, from the $(p-1)$ possibilities, is at most $\frac{(p-1)}{m}$. So, the number of bad hash functions $h \in \mathcal{H}$ is equal to $p\frac{(p-1)}{m} = \frac{|H|}{m}$, in the end:

$$Pr(\{h \in \mathcal{H} \mid h(k_1) = h(k_2)\}) = \frac{\#\text{ bad choices of } h}{\#\text{ all choices of } h \in \mathcal{H}} = \frac{|\mathcal{B}|}{|\mathcal{H}|} = \frac{\frac{|\mathcal{H}|}{m}}{|\mathcal{H}|} = \frac{1}{m}$$

$\mathcal{H}$ is an *universal* hash family.

□