Universality of Hashing

[Notes for the Training Camp]

Yufei Tao

ITEE
University of Queensland

In this lecture, we will prove the universality of the hash function we designed in the main class. Our proof serves as a nice illustration of how computer science can benefit from number theory.

Recall:

Hash Function

Let U and m be positive integers.

A hash function is a function h that maps $[U]$ to $[m]$ (recall that $[x]$ denotes the set of integers $\{1, 2, ..., x\}$), namely, for any integer $k \in [U]$, $h(k)$ returns a value in $[m]$.

Recall:

Universality

Let $\mathcal{H}$ be a family of hash functions. $\mathcal{H}$ is universal if the following holds:

> Let k_1, k_2 be two distinct integers in $[U]$. By picking a function $h \in \mathcal{H}$ uniformly at random, we guarantee that
>
> $$\boldsymbol{Pr}[h(k_1) = h(k_2)] \quad \leq \quad 1/m.$$

Recall:

A Universal Family

Pick a prime number p such that $p \geq \max\{U, m\}$. Choose an integer α uniformly at random from $\{1, 2, ..., p-1\}$, and an integer β uniformly at random from $\{0, 1, ..., p-1\}$. Design a hash function as:

$$h(k) \quad = \quad 1 + ((\alpha \cdot k + \beta) \mod p) \mod m$$

The Prime Ring

Denote by $\mathbb{Z}_p$ the set of integers $\{0, 1, ..., p-1\}$. $\mathbb{Z}_p$ forms a commutative ring under $+$ and $\cdot$ modulo p. This means:

- $\mathbb{Z}_p$ is closed under $+$ and $\cdot$ modulo p.
- $+$ modulo p satisfies commutativity and associativity.
  - $a + b = b + a \pmod p$ and $a + b + c = a + (b + c) \pmod p$
- $+$ modulo p has a zero element, that is, $0 + a = a \pmod p$.
- Every element a has an additive inverse $-a$, that is, $a + (-a) = 0 \pmod p$.
- $\cdot$ modulo p satisfies commutativity and associativity.
  - $a \cdot b = b \cdot a \pmod p$ and $a \cdot b \cdot c = a \cdot (b \cdot c) \pmod p$
- $\cdot$ modulo p has a one element, that is, $1 \cdot a = a \pmod a$.
- $+$ and $\cdot$ modulo p satisfy distributivity.
  - $a \cdot (b + c) = a \cdot b + a \cdot c \pmod p$
  - $(b + c) \cdot a = b \cdot a + c \cdot a \pmod p$

The Prime Ring

The ring $\mathbb{Z}_p$ has several crucial properties. Let us start with:

Lemma: Let a be a non-zero element in $\mathbb{Z}_p$. Then, $a \cdot j \neq a \cdot k \pmod p$ for any $j, k \in \mathbb{Z}_p$ with $j \neq k$.

Proof: Suppose without loss of generality $j > k$. Assume $a \cdot j = a \cdot k \pmod p$, then $a \cdot (j - k) = 0 \pmod p$. This means that $a \cdot (j - k)$ must be a multiple of p. Since p is prime, either a or $j - k$ must be a multiple of p. This is impossible because a and $j - k$ are non-zero elements in $\mathbb{Z}_p$. □

The lemma implies that $a \cdot 0$, $a \cdot 1$, ..., $a \cdot (p - 1)$ must take unique values in $\{0, 1, ..., p - 1\}$.

The Prime Ring

The previous lemma immediately implies:

Corollary: Every non-zero element a has a unique multiplicative inverse a^{-1}, namely, $a \cdot a^{-1} = 1 \pmod{p}$.

In other words, $\mathbb{Z}_p$ is a division ring.

The Prime Ring

The next property then follows:

Lemma: Every equation $a \cdot x + b = c \pmod{p}$ where a, b, c are in $\mathbb{Z}_p$ and $a \neq 0$ has a unique solution in $\mathbb{Z}_p$.

Proof:

$$\begin{aligned} a \cdot x &= c - b \pmod{p} \\ \Rightarrow \quad x &= a^{-1} \cdot (c - b) \pmod{p} \end{aligned}$$

□

Proof of Universality

Next, we will prove that the hash family $\mathcal{H}$ defined in Slide 5 is universal. As before, let k_1 and k_2 be distinct integers in $[U]$.

Fact 1: Let

$$\begin{aligned} g(k_1) &= (\alpha \cdot k_1 + \beta) \mod p \\ g(k_2) &= (\alpha \cdot k_2 + \beta) \mod p \end{aligned}$$

Then, $g(k_1) \neq g(k_2)$.

Proof: Otherwise, it must hold that

$$\begin{aligned} \alpha \cdot k_1 + \beta &= \alpha \cdot k_2 + \beta \pmod p \\ \Rightarrow \quad \alpha \cdot (k_1 - k_2) &= 0 \pmod p \end{aligned}$$

which is not possible. $\square$.

Proof of Universality

How many different choices are there for the pair $(g(k_1), g(k_2))$? The answer is at most $p(p-1)$ according to Fact 1 – there are p^2 possible pairs in $\mathbb{Z}_p \times \mathbb{Z}_p$ but we need to exclude the p pairs where the two values are the same.

How many different hash functions are there in $\mathcal{H}$? The answer is obviously $p(p-1)$ because there are $p-1$ selections for α, and p selections for β.

Next, we will prove a one-to-one mapping between the possible choices of $(g(k_1), g(k_2))$ and the hash functions in $\mathcal{H}$.

Proof of Universality

Fact 2: Fix any two $x, y \in \mathbb{Z}_p$ such that $x \neq y$. There is a unique hash function $h \in \mathcal{H}$ such that $h(k_1) = x$ and $h(k_2) = y$.

Proof: Suppose that h is determined by α, β selected as explained in Slide 5. Thus:

$$\begin{aligned} \alpha \cdot k_1 + \beta &= x \pmod{p} \\ \alpha \cdot k_2 + \beta &= y \pmod{p} \end{aligned}$$

Hence:

$$\begin{aligned} \alpha \cdot (k_1 - k_2) &= x - y \pmod{p} \\ \Rightarrow \quad \alpha &= (k_1 - k_2)^{-1} \cdot (x - y) \pmod{p} \\ \Rightarrow \quad \beta &= x - (k_1 - k_2)^{-1} \cdot (x - y) \cdot k_1 \pmod{p} \end{aligned}$$

□

Proof of Universality

Let P be the set of pairs (x, y) such that $x, y \in \mathbb{Z}_p$ and $x \neq y$.

We know that by choosing $h \in \mathcal{H}$ randomly, we are essentially picking a pair (x, y) for $(g(k_1), g(k_2))$ uniformly at random.

Notice that $h(k_1) = h(k_2)$ if and only if $g(k_1) = g(k_2) \pmod{m}$. So now the question boils down to: how many pairs (x, y) in P satisfy $x = y \pmod{m}$?

Proof of Universality

How many pairs (x, y) in P satisfy $x = y \pmod{m}$?

- For $x = 0$, y can take $m, 2m, 3m, \ldots$ – definitely no more that $\lceil p/m \rceil - 1 \leq (p-1)/m$ choices
- For $x = 1$, y can take $m+1, 2m+1, 3m+1, \ldots$ – definitely no more that $\lceil p/m \rceil - 1 \leq (p-1)/m$ choices
- ...

Hence, the number of such pairs is no more than $p(p-1)/m = |P|/m$.

Now we conclude that the probability of $h(k_1) = h(k_2)$ is at most $1/m$.