Tutorial 11

17 May 2023 14:08

Fingerprinting II and string comparison

Schwartz-Zippel thm

$$\Pr\left(Q(r_1,\dots,r_n)=0 \mid Q\neq 0\right) \leq \frac{\deg Q}{|S|}$$

$r_i \in_R S$

Problem Verify whether two strings X and Y $X, Y \in \{0,1\}^n$ are identical.

Deterministically $O(n)$

$X = (x_1,\dots,x_n)$

$Y = (y_1,\dots,y_n)$ $\quad x_i, y_i \in \{0,1\}$

If comparison is an expensive operation, then SZT gives us the following solution:

- Interpret X and Y as polynomials (variable)

$$X(z_1,\dots,z_n) = \sum_{i=1}^{n} x_i z_i \mod p$$

$$Y(z_1,\dots,z_n) = \sum_{i=1}^{n} y_i z_i \mod p$$

$$X(\vec{z}) - Y(\vec{z}) \stackrel{?}{\neq} 0$$

Choose $\vec{r} \in_R \{0,1\}^n$

by SZT

$$\Pr\left(X(\vec{r}) - Y(\vec{r}) = 0 \mid [X-Y][\vec{z}] \neq 0\right) \leq \frac{\deg(X-Y)}{2} = \frac{1}{2}$$

Context: Database comparisons

→ Two distant databases X and Y. Are they the same?

→ Expensive operation: transmitting a bit

Is the method above efficient?

NO. Random r needs to be sent and it is as large as the database

Solution 1

Interpret X and $\vec{i}$ as numbers:

$$num(X) = \sum_{i=1}^{n} x_i 2^{i-1}$$

$$num(\vec{i}) = \sum_{i=1}^{n} b_i 2^{i-1}$$

Compare

$X \bmod p$ and $Y \bmod p$ (fingerprints)

for a suitably randomly chosen prime p. Specifically $p < k$.

If p is small, fingerprints are also small. However there is a trade-off between the size of p and the probability of an error.

Error can happen if $X \neq Y$ but $X \equiv Y \pmod p$

$x - y \equiv 0 \mod p$ ($x-y$ is divisible by p)

$\pi(k)$ - number of primes smaller than k.

$\pi(k) = \Theta\left(\frac{k}{\ln k}\right)$

for $k \geq 29$ $\pi(k) \geq 1.2 \cdot \frac{k}{\ln k}$ &

$$Pr(x - y = 0 \mod p \mid x \neq y) = \frac{\#\text{ bad primes}}{\#\text{ primes we choose from}} \leq \frac{n}{\pi(k)} \leq \frac{\ln k \cdot n}{1.2 \cdot k}$$

bad primes: How many primes can divide $x-y$?

$x - y < 2^n$

What is the smallest number with n distinct prime divisors?

$$\prod_{i=1}^{n} p_i > 2^n = \prod_{i=1}^{n} 2$$

p_i - i^{th} smallest prime

for $k = t \cdot n \ln(t \cdot n)$

$$Pr < \frac{\ln(t \cdot n \ln(t \cdot n)) \cdot n}{t \cdot n \ln(tn)} \in O\left(\frac{1}{t}\right)$$

How many bits x needs to send to y?

for $t = n$ a prime of $\log(k)$ bits $k = n^2 \cdot 2 \ln(n)$

$\in O(\log n)$

and the finger print of the same size.

$$X = \sum_{i=1}^{n} x_i z^{i-1} \mod p$$

Solution 1

Choose $z = 2$ and randomize over p

Solution 2

Choose p and randomize over z

$$v \in_R S \subseteq \mathbb{Z}_p$$

$$Pr\left[(X-Y)(v) = 0 \bmod p \mid (X-Y)(z) \neq 0 \right] \leq \frac{\deg(X-Y)}{|S|} = \frac{n-1}{|S|}$$

To compare with solution 1, we would like the probability of success to be roughly $\frac{1}{n}$.

$\Rightarrow |S| = n^2 \Rightarrow$ p needs to be larger than n^2

What needs to be sent?

$v < p \approx O(\log(n))$ bits

and $X(v) \bmod p$ $O(\log(n))$ bits

3rd method

Choose a random polynomial P mod p and evaluate

$P(\text{num}(X))$ and $P(\text{num}(Y))$ and compare

$\Rightarrow$ UNIVERSAL HASHING

$\mathcal{H}: \quad X \to Y \qquad c: \Pr(f(H) = c) = \frac{\mathcal{H}}{|Y|}$