Tutorial 13

27 May 2021 16:05

Finger printing and string comparison

Schwartz-Zippel theorem

$$\Pr\left(Q(r_1,\dots,r_n)=0 \mid Q \not\equiv 0\right) \le \frac{\deg Q}{|S|}$$

$r_i \in_R S$

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

$\boxed{\text{Deterministically} \quad O(n)}$ $X = (x_1,\dots,x_n)$ $x_i, y_i \in \{0,1\}$

$Y = (y_1,\dots,y_n)$

If comparison is an expensive operation, then S-Z theorem gives us solution:

→ Interpret X and Y as polynomials:

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

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

$$X(\vec{z}) - Y(\vec{z}) \overset{?}{\equiv} 0$$

$\boxed{\text{Choose } \vec{v} \in \{0,1\}^n}$

by S-Z theorem

$$\Pr\left[X(\vec{r}) - Y(\vec{r}) = 0 \mid (X-Y)[\vec{z}] \not\equiv 0\right] \le \frac{\deg(X-Y)}{2} = \frac{1}{2}$$

Context: Database comparison

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

→ Expensive operation: transmitting a bit (sending messages between the databases).

Is the method above efficient? ↑

No! Random r needs to be distributed and it is as long as the database!

Solution 1

Interpret both X and Y as numbers:

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

$$num(Y) = \sum_{i=1}^{n} y_i \cdot 2^{i-1}$$

Compare

$\underline{X \bmod p}$ and $\underline{Y \bmod p}$ fingerprints

for suitably + randomly chosen prime p. (smaller than k) If p is small, fingerprints are small (in bits). However there is a tradeoff between the size of p and the probability of an error.

Error can happen if $X \not\equiv Y$ but $X \equiv Y \bmod p$

$\boxed{X - Y \equiv 0 \mod p}$ (read X-Y is divisible by p)

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

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

for $k \geq 29$ $\pi(k) \leq 1.2 \dots \cdot \dfrac{k}{\ln k}$

for $k \geq 29$ $\quad \pi(k) \leq 1.2 \ldots \cdot \frac{k}{\ln k}$

$$\Pr\left(X - Y \equiv 0 \mod p \mid X \neq Y\right) = \frac{\#\text{ bad primes}}{\#\text{ primes we chose from}} = \frac{n}{\pi(k)} < \frac{\ln k \cdot n}{k \cdot (1.2)}$$

\# bad primes: How many prime divisors can $X-Y$ have at most?

What is the largest value of $X-Y$? $\quad X-Y < 2^n$

What is smallest number with $\underline{n}$ prime divisors?

$$\prod_{i=1}^{n} p_i > 2^n = \prod_{i=1}^{n} 2 \Rightarrow \#\text{ bad primes} < n$$

p_i - i^{th} smallest prime

for $\boxed{k = t \cdot n \ln(tn)}$

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

How many bits does X need to send to Y?

for $\boxed{t = n}$ a prime of $O(\log_2(n))$ bits

and the finger print $O(\log_2 n)$

$$\boxed{X = \sum_{i=0}^{n} x_i \cdot z^{i-1} \mod p}$$

Solution 1:

Choose $z = 2$ and randomize over p.

Solution 2:

Choose p and randomize over z

Method 2 analysis using S-Z theorem

$X(z)$ and $Y(z)$ are polynomials mod p

$r \in_R S \subseteq \mathbb{Z}/p$

$$\Pr\left[(X-Y)(r) = 0 \mod p \mid (X-Y)(z) \not\equiv 0\right] \leq \frac{\deg(X-Y)}{|S|} = \frac{n-1}{|S|}$$

To match the method 1, we would like this probability to be roughly $1/n$

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

What needs to be sent?

$$\left.\begin{array}{ll} r < p \approx & O(\log(n)) \text{ bits} \\ \text{and } X(r) \mod p & O((\log(n))) \text{ bits} \end{array}\right\} \text{twice as large}$$

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