HONORS THEORY OF ALGORITHMS

2024-01-05

ANDREW CHI-CHIH YAO 1979
continuous version → HAROLD ABELSON 1978

COMMUNICATION COMPLEXITY

Alice access to string X length $|X|=N$
Bob Y $|Y|=N$

cooperatively evaluate $f(X,Y)$

trivial algorithm: N bits comm.

$$(\forall f)(CC(f) \leq N)$$

$$Id(X,Y) = \begin{cases} 1 & \text{if } X=Y \\ 0 & \text{if } X \neq Y \end{cases}$$

Communication matrix

$$M_f = \left(f(x,y)\right)_{2^N \times 2^N}$$

Mehlhorn - Schmidt

Thm $\quad CC(M_f) \geq \log_2 \operatorname{rk} M_f$

$\therefore \quad CC(Id_N) = N$

RANDOMIZED CC

Rabin - Yao - Simon RYS protocol for Id

$N = |X|, |Y|$, k: $2 \le k << N$

1. Alice generates random prime $p < 2^k$ — k bits
2. Alice computes $(X \bmod p)$ — -"-
3. A → B: p, $(X \bmod p)$ — 2k bits of comm.
4. Bob computes $(Y \bmod p)$
5a. If $(X \bmod p) \neq (Y \bmod p)$ Bob says "$X \neq Y$"
5b. Else — -"- "$X = Y$"

Analysis: ~~Correctness~~ Cost — $Pr(\text{error}) < ?$

$N = 1 \text{ petabyte} \approx 8 \cdot 10^{15} \text{ bits} = 2^{53} \text{ bits}$

$k = \sout{150}\ 200$

$Pr < 10^{-41}$

$X \neq Y$

$$P(\text{error}) = \frac{\#\text{distinct primes } p \le 2^k \text{ s.t. } p \mid X-Y}{\pi(2^k)} = \frac{\text{num}}{\text{den}} <$$

case of error:

$p \mid X-Y$

$\uparrow$ divides

$X \equiv Y \bmod p$

$\#\text{primes} \le 2^k$ $\uparrow$

$$< \frac{N}{\frac{2^k}{\ln 2^k}} < \frac{k \cdot N}{2^k}$$

PRIME NUMBER THM

$$\pi(x) \sim \frac{x}{\ln x} \qquad 1896$$

meaning

$$\lim_{x \to \infty} \frac{\pi(x)}{\frac{x}{\ln x}} = 1$$

eqns thm

num $<$ #distinct primes $p \mid X-Y \le N$

$X-Y$ ~ N-bit number

$$2 \cdot \frac{200 \cdot 8 \cdot 10^{15}}{2^{200}} \approx \frac{3.2 \cdot 10^{18}}{10^{60}} = 3.2 \cdot 10^{-42} < 10^{-41}$$

$$\pi(x) > \frac{1}{2} \cdot \frac{x}{\ln x}$$ ← Chebyshev ~1850

for $x \ge x_0$

A: knows X, p

B: " $Y, p, (X \bmod p)$

both know: $X \neq Y \bmod p$

HW: deterministically find $i \in [N]$ s.t. $X_i \neq Y_i$

$[N] = \{1, \ldots, N\}$

with "small" amount of communication

$\text{poly}(k, \log N)$ e.g. $k^2 \log N$

polynomial of ...

ASY, DM mini ← online notes

Asymptotic notation: big-Oh

a_n, b_n sequences of reals

We say that

$$a_n = O(b_n)$$

"is"

if

$$(\exists C)(\text{for all sufficiently large } n)(|a_n| \le C|b_n|)$$

$(\exists n_0)(\forall n)(\text{if } n \ge n_0 \text{ then } \ldots)$

threshold

Example:

$$1000 \cdot x^5 + 7x^3 + 1500 = O(x^5)$$

upper bound