ÇUKUROVA UNIVERSITY
FACULTY OF ENGINEERING
COMPUTER ENGINEERING DEPARTMENT

Lecture 7: Probabilistic (or Randomized) Algorithms

Example 1: Dataset Comparison

$$X = x_1 x_2 \ldots x_n, \qquad x_i \in \{0,1\}$$

$$Y = y_1 y_2 \ldots y_n, \qquad y_i \in \{0,1\}$$

Question: $X \stackrel{?}{=} Y$

Classical Communication Protocol:

Comparing the two datasets bit by bit:

- Communication overhead: $O(n)$
- Comparison overhead: $O(n)$ => $n = 10^{16} bits\ (B) \approx 1136\ TB$

$$v(X) = value(X) = \sum_{1}^{n} x_i \, . \, 2^{n-i}$$

$$\Pi(k) = \{p \in \mathbb{P} \mid p \leq k\}, \ \pi(k) = |\Pi(k)|$$

Randomized Equal algorithm:

$$Input: \quad \underset{A}{X = x_1 x_2 \ldots x_n}, \quad \underset{B}{Y = y_1 y_2 \ldots y_n}, \quad x_i y_i \in \{0,1\}$$

1. A randomly chooses a prime number $p \in \Pi(n^2)$
2. A calculates the "fingerprint of X => s = v (X) mod p.
   a. **Note:** X is considered as a binary number
3. A sends s and p to B
4. B calculates the "fingerprint of Y => t = v (Y) mod p
5. B compares whether s = t? Are the two fingerprints the same?
   a. Yes: equal -> A,
   b. No: unequal -> A

Comparison overhead:

The Equal algorithm is a randomized communication protocol for the data comparison example. It drastically reduces the comparison effort. Comparison overhead was previously $O(n)$, now the overhead is reduced to 5 $steps \rightarrow O(1)$

Communication overhead:

$$0 \leq p, s \leq n^2$$
$$m \in \mathbb{N} \rightarrow \lceil \log m \rceil \; Bits$$
$$l(s,p) \leq 2 * \lceil \log n^2 \rceil \leq 4 * \lceil \log n \rceil$$

For $n = 10^{16}$;

$$\leq 4 * 16 * \lceil \log 10 \rceil$$
$$\leq 4 * 16 * 4$$
$$= 256 \text{ Bits}$$

Instead of 10^{16}, we would only transmit 256 bits.

$$Prob_{equal}[A \mid B]$$

$$X = Y => t = s$$
$$X \neq Y <= t \neq s$$

$$Prob_{equal}["unequal" \mid X = Y]) = 0$$

Example:

$$n = 5$$

$$X = 10011 \quad v(X) = 19$$
$$Y = 10001 \quad v(Y) = 17$$

$$for\; p = 11 => \text{ s } = \text{ 8, t } = \text{ 6}$$

$$X = 10011 \quad v(X) = 19$$
$$Y = 10001 \quad v(Y) = 17$$

$$for\; p = 2 => \text{ s } = \text{ 1, t } = \text{ 1}$$

<u>$p = 2$ is a bad witness for the inequality of X and Y</u>

Question: How many bad witnesses are in $\Pi(n^2)$?

$s = t$, i.e. $v(X) = v(X)\; mod\; (p), although\; X \neq Y.$

$$\Pi^{+}(n^2, X, Y) = \{p \in \mathbb{P} \mid v(X) \neq v(Y)(p), X \neq Y\}$$

$$\Pi^{-}(n^2, X, Y) = \{p \in \mathbb{P} \mid v(X) = v(Y)(p), X \neq Y\}$$

Establishing the relationship between bad witnesses and total witness candidates:

$$Prob_{equal}[X = Y \mid X \neq Y]) = \frac{|\Pi^{-}(n^2, X, Y)|}{|\Pi(n^2)|} = \frac{\mathbf{2 \ln n}}{\boldsymbol{n}}$$

$$for\ n = 10^{16} \ => \frac{2 \ln 10^{16}}{10^{16}} \approx 0.7 * 10^{-14}$$

Multiple rounds of algorithm to further reduce the error:

$$l = 10 => 0.7 * 10^{-144}$$

$$O(l * \log n)$$

$$L \subseteq \Sigma^*, w \in \Sigma^*: w \in^{?} L$$

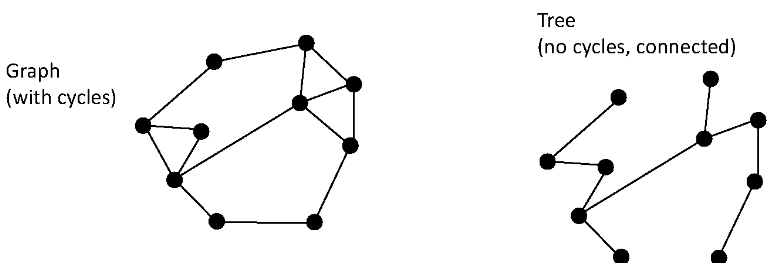
$$D = \{(x, y) \in \{0,1\}^n x \{0,1\}^n \mid n \in \mathrm{N}_{0,}, x \neq y\}$$

Given $w \in \{0,1\}^n x \{0,1\}^n$

1. $Prob_{equal}["w \in D" \mid w \notin D] = 0$
2. $Prob_{equal}["w \in D" \mid w \in D] \geq 1 - \frac{2 \ln n}{n} \quad => RP(\varepsilon(n)),\ \varepsilon(n)\ is\ the\ error\ bound$
3. Overhead: $O(\log n)$

Example 2: Triangle Graph

$$\Delta Graph = \{\langle G \rangle \mid G\ undirected\ graph\ that\ contains\ at\ least\ one\ triangle\}$$



Randomized Triangle Graph Algorithm:

1. $T\ (tester)$ randomly chooses an edge $\{a, b\}$
2. T randomly chooses a node $c \neq \{a, b\}$
3. Test: Do c and $\{a, b\}$ form a triangle? -> Yes or No

$$Prob_T["yes" \mid G \notin \Delta Graph] = 0$$ ➔ no false positive statement

$$G = (V, E)$$

$$|V| = n, |E| = m$$

$$Prob_T["yes\ in\ one\ round" \mid G \in \Delta Graph] \geq \frac{3}{m} * \frac{1}{n-2}$$

$$Prob_T["no\ in\ one\ round" \mid G \in \Delta Graph] \leq 1 - \frac{3}{m(n-2)}$$

$$Prob_T["no\ in\ l\ rounds" \mid G \in \Delta Graph] \leq \left(1 - \frac{3}{m(n-2)}\right)^l$$

$$Prob_T["yes\ in\ l\ rounds" \mid G \in \Delta Graph] \geq 1 - \left(1 - \frac{3}{m(n-2)}\right)^l$$

$$\left(1 + \frac{1}{k}\right)^k \xrightarrow{k \to \infty} e^1 = 2.7182 \ldots$$

More general formulation with x instead of the special case with 1:

$$\left(1 + \frac{x}{k}\right)^k \xrightarrow{k \to \infty} e^x$$

Place a minus before x:

$$\left(1 + \frac{-x}{k}\right)^k \xrightarrow{k \to \infty} e^{-x}$$

We want to apply the Euler sequence for our probabilistic $\Delta Graph$ algorithm:

$$k = l, x = yl \ ➔\ (1 - y)^l \xrightarrow{l \to \infty} e^{-yl}$$

$$y = \frac{3}{m(n-2)}, l = \frac{m(n-2)}{3} \ ➔\ y * l = 1$$

$$\left(1 - \frac{3}{m(n-2)}\right)^{\frac{m(n-2)}{3}} \approx e^{-1} \approx \frac{1}{e} \approx \frac{1}{2.7} < \frac{1}{2}$$

RP

1. $Prob_T["yes" \mid G \notin \Delta Graph] = 0$
2. $Prob_T["yes" \mid G \in \Delta Graph] \geq \frac{1}{2}$
3. $polynomiel$

Example 3: Probabilistic Prime Number Tests

Little Fermat's Theorem:

$$p \in \mathbb{P}, a \in \mathbb{N}, (a, p) = 1$$
$$\Downarrow$$
$$a^{p-1} = 1(p)$$

Question: $n \in \mathbb{N}, a \in \mathbb{N}\ with\ a^{n-1} = 1(n) => n \overset{?}{\in} \mathbb{P}$

$$2^{n-1} = 1(n)$$
$$n = 3, 4, 5 \ldots$$

n=341 – the number is not prime, but it pretends to be prime

⇨ Smallest pseudoprime number to base 2

Definition: Let m be a composite number with $(a, m) = 1$ and $a^{m-1} = 1(m)$ or $a^m = a(m)$, then m is called pseudoprime to base a

n=341 – the number is not prime, but it pretends to be prime

- n=341 is smallest pseudoprime number to base 2
- n=341 is not a pseudoprime to base 3

In other words: Pseudoprime numbers satisfy Fermat's little theorem even though they are not prime

Existence of composite numbers that are pseudoprime to all (coprime) bases:

A composite number is called a Charmichael number iff $a^{m-1} = 1(m)$ or $a^m = 1(m)$ applies to all bases with $(a, m) = 1$

- m=561 is the smallest Carmichael number
- Carmichael numbers are free of squares
- Factoring Carmichael numbers contains at least 3 different prime factors
- There are infinitely many Carmichael numbers: 561, 1105, 1729, 2465, 2821, ...

Theorem: For $m \in \mathbb{N}, m \geq 3$, let

$$F_m = \{a \in \mathbb{Z}_m \mid a^{m-1} = 1(m)\}$$

be the set of bases for which m passes the Fermat test.

If m is not a prime number, then F_m contains the bases that "fool" the Fermat test.

Let $m \in \mathbb{N}, m \geq 3$, be a composite and not a Carmichael number, then the following applies:

$$|F_m| \leq \frac{\mathbb{Z}_m}{2}$$

```
alg notPrime (k ∈ U+, k ≥ 3)

      Randomly pick an a ∈ (1, …, k − 1) with (a, k) = 1

      If a^(k−1) ≠ 1(k)
            then Output:  k is not prime
            otherwise Output:  k prime?
endalg
```

$\overline{\mathbb{P}}$ = COMPOSITES $\in$ RPP

(1) $Prob[k \in \overline{\mathbb{P}} \mid k \notin \overline{\mathbb{P}}] = 0$

(2) $Prob[k \notin \overline{\mathbb{P}} \mid k \in \overline{\mathbb{P}}] \leq \frac{1}{2}$

(3) notPrime is polynomial

The probability of error is therefore at most $\frac{1}{2}$. If the algorithm is now carried out for l rounds in which the base a is chosen anew at random and independently, then the probability of error is at most $\frac{1}{2^l}$; so, it can be made as small as you want.

➔ Executing the algorithm l times leads to an error probability of $\leq \frac{1}{2^l}$

Miller Rabin Algorithm and b-Sequences

$$\mathbb{Z}_m \text{ is a field iff } m \in \mathbb{P}$$

$$x^2 = 1$$

$$\mathbb{Z}_m: x = 1, x = -1 \implies \text{trivial solutions}$$

$x \in \mathbb{Z}_m$ is called a non-trivial square root of 1 modulo m if $x^2 = 1$ and $x \neq 1$ and $x \neq -1$.

e.g., $m = 15$: $x = 4, x = -4 \implies non-trivial\ solutions$

In $\mathbb{Z}_m$, x^2 has the solution $x = \pm 1$ iff $m \in \mathbb{P}$

In other words: If there is a nontrivial square root modulo m, then m is a composite number.

$$m \in \mathbb{N}$$

$$s = max\{\, r \in \mathbb{N} \mid 2^r \mid m-1\}$$

$$d = \frac{m-1}{2^s}$$

$$b \in \mathbb{Z}_m: b-sequence$$

$$\langle b^{2^0 d}, b^{2^1 d}, b^{2^2 d}, \dots, b^{2^{s-1} d}, b^{2^s d} \rangle \ \ mod\ \ (m)$$

If $m \in \mathbb{P}$:

$$b^{2^s d} = b^{2^s * \frac{m-1}{2^s}} = b^{m-1} = 1(m)$$

➔ **We get more information about the structure of the b-sequence if m is a prime number:**

Examples:

a) Let m=25 (i.e., composite), then s=3 and d=3. For the basis b=2 the sequence results

$$\langle 2^3, 2^6, 2^{12}, 2^{24} \rangle = \langle 8,14,21,16 \rangle$$

For b=3 we get

$$\langle 3^3, 3^6, 3^{12}, 3^{24} \rangle = \langle 2,4,16,6 \rangle$$

and for $b = 7$

$$\langle 7^3, 7^6, 7^{12}, 7^{24} \rangle = \langle 18, -1,1,1 \rangle$$

b) Let m=97 (i.e., prime), then s=5 and d=3. For the basis b=2 the sequence results

$$\langle 2^3, 2^6, 2^{12}, 2^{24}, 2^{48}, 2^{96} \rangle = \langle 8,64,22, -1,1,1 \rangle$$

For b=14 this results

$$\langle 14^3, 14^6, 14^{12}, 14^{24}, 14^{48}, 14^{96} \rangle = \langle 28,8,64,22, -1,1 \rangle$$

for $b = 35$

$$\langle 35^3, 35^6, 35^{12}, 35^{24}, 35^{48}, 35^{96} \rangle = \langle 1,1,1,1,1,1 \rangle$$

and for $b = 62$

$$\langle 62^3, 62^6, 62^{12}, 62^{24}, 62^{48}, 62^{96} \rangle = \langle -1,1,1,1,1,1 \rangle$$

➔ The examples show that the b-sequences for prime numbers have a specific structure.

Theorem: *Let be* $p \in \mathbb{P}$, $s = max\{r \mid 2^r \mid p-1\}$, $d = \frac{p-1}{2^s}$, $b \in \mathbb{N}$ *with* $(b,p) = 1$

Then (1) $b^d = 1\ (p)$ or

(2) $\exists r \in \{0, 1, \dots, s-1\}: b^{2^r d} = -1\ (p)$

If $p \in \mathbb{P}$, then the b-sequence has one of the following forms:

(1) $\langle 1,\ 1, \dots, 1 \rangle$

(2) $\langle -1,\ 1, \dots\ ,\ 1 \rangle$

(3) $\langle ?,\ ?,\ \dots\ ,\ ?, -1, 1, \dots, 1 \rangle$

Last element of the b-sequence: $b^{p-1} = 1\ (p)$

Use the reverse of the b-sequence:

$m \in \mathbb{N}, b \in \mathbb{Z}_m, with\ b - sequence$

(1) $\langle ?,\ ?, \dots ,1,1, \dots ,1\rangle$

(2) $\langle ?, ?, \dots, ?, -1\rangle$

(3) $\langle ?,\ ?,\ \dots\ ,\ ?, ?\rangle$

➔ $m \notin \mathbb{P}$

Definition: $m \in U_+, m \geq 3, m - 1 = 2^s d\ \ with\ \ d \in U_+, b \in \mathbb{Z}_m$

If $b^d = 1(m)$ or $b^{2^r d} = -1(m)$ holds for an $r \in \{0, 1, \dots, s - 1\}$, then m is called **strong pseudoprime** to base b.

Theorem: $m \in U_+, m \geq 3, composite$, then the number of bases for which m is **strongly pseudoprime** is at most $\frac{\boldsymbol{m-1}}{\boldsymbol{4}}$

➔ Miller Rabin Algorithm

```
algorithm MILLER-RABIN(n ∈ U+, n ≥ 3)
    Compute d and s with n − 1 = d · 2^s and d positive uneven
    Randomly pick an a ∈ {2, 3, . . . , n − 2}
    b := a^d(n)
    if b = 1 (n) or b = −1 (n): Output: n is prime?
    for r := 1 to s − 1 do
        b := b^2 (n)
        if b = −1 (n): Output: n is prime?
        if b = 1 (n): Output: n is not prime
    endfor
    Output: n is prime?
endalgorithm MILLER-RABIN
```

Miller Rabin Algorithm ∈ RPP

(1) $Prob[Output\!: n\ is\ not\ prime \mid n \in \mathbb{P}] = 0$

(2) $Prob[Output\!: n\ is\ prime? \mid n \notin \mathbb{P}] \leq \frac{1}{4}$

(3) $Prob[l - times\ Output\!: n\ is\ prime? \mid n \notin \mathbb{P}] \leq \frac{1}{4^l}$

(4) $Prob[after\ l\ excecutions\ Output\!: n\ is\ not\ prime \mid n \notin \mathbb{P}] \geq 1 - \frac{1}{4^l}$

(5) $O\big(l * (\log n)\big)$ arithmetical operations or $O(l * (\log n)^3)$ bit operations

Outline:

$Dataset\ \in RP\left(\frac{2\ln n}{n}\right)$

$\Delta Graph \in RP\left(\frac{1}{2}\right)$

$Miller\ Rabin \in RP\left(\frac{1}{4}\right)$

Complexity Class RP:

Random/Probabilistic polynomial running algorithms with one-sided error

- No false positive statements.
- Conversely, RP algorithms make errors with a particular bound ε.
- By repeating the execution of the algorithm, the total error can be reduced.

⇨ The algorithms that class RP defines are also called **Monte Carlo algorithms**. Such algorithms allow a one-sided error.