CSE 312

Foundations of Computing II

Lecture 28: Randomized Algorithms II

PAUL G. ALLEN SCHOOL
OF COMPUTER SCIENCE & ENGINEERING

Stefano Tessaro

tessaro@cs.washington.edu

Today

- Randomized algorithms: Polynomial-identity testing
  - An Application: Hashing!
- Wrap-up
- Also: There are office hours today!
  - Leo & Siva will each hold one hour!
  - There will be office hours on Monday
  - Stay tuned!

PLEASE

COMPLETE

THE

CLASS

EVALUATION

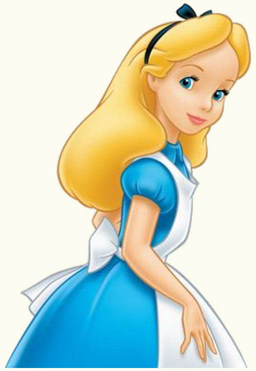
!!!!!!!

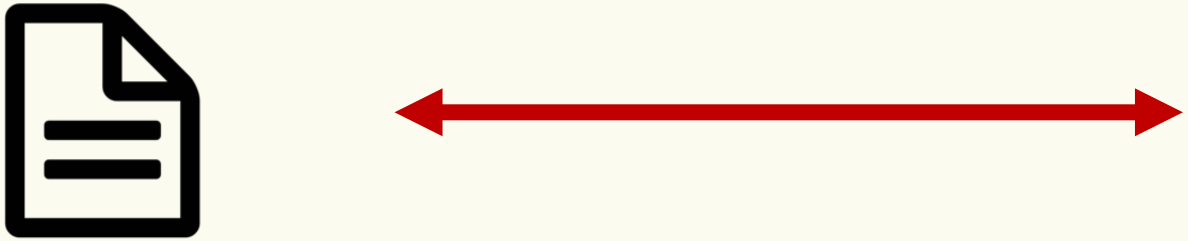
Problem

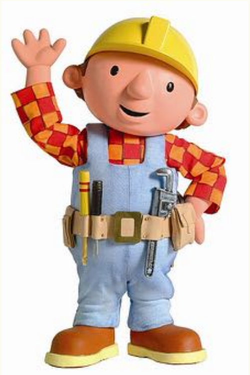
Goal: Alice and Bob want to know whether they have the same file, by **communicating as little as possible.**

1 GB file A

1 GB file B







If they want to be absolutely certain, they need to communicate 1GB of data in the worst case.

What if they accept some small error probability? (Say at most 1/16?)

We will see: Answer approx 64 bits = 8 bytes!

Polynomials

Definition. A **polynomial** is a formal expression of the form $a(X) = a_0 + a_1X + a_2X^2 + \cdots + a_dX^d$, where $a_0, a_1, \ldots, a_d$ are the numbers (the **coefficients**) and d is the **degree**.

Examples:

- $1 + X + X^2$
- $1 + 3X + 5X^5$

Polynomials modulo a prime

We denote $\mathbb{Z}_p = \{0,1,\dots,p-1\}$

Definition. A **polynomial** mod p is a formal expression of the form $a(X) = a_0 + a_1X + a_2X^2 + \cdots + a_dX^d$, where $a_0, a_1, \dots, a_d \in \mathbb{Z}_p$ are the coefficients and d is the degree.

Polynomials modulo a prime

Definition. A **polynomial** mod a prime p is a formal expression of the form $a(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_d X^d$, where $a_0, a_1, \ldots, a_d \in \mathbb{Z}_p$ are the coefficients and d is the degree.

Definition. The **evaluation** of $a(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_d X^d$ at $x \in \mathbb{Z}_p$ is the value

$$a(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \bmod p$$

Example

$p = 7, a(X) = 1 + X + X^2$

- $a(0) = 1$
- $a(1) = 1 + 1 + 1 = 3$
- $a(2) = 1 + 2 + 4 = 7 \bmod 7 = 0$
- $a(3) = 1 + 3 + 2 = 6$
- $a(4) = 1 + 4 + 2 = 0$
- $a(5) = 1 + 5 + 4 = 3$
- $a(6) = 1 + 6 + 1 = 1$

Here, 2 and 4 are the **zeros** of $a(X)$

Zeros of Polynomial

Non-zero = at least one coefficient is not zero!

Q: How many zeros does a polynomial $a(X)$ mod p have?

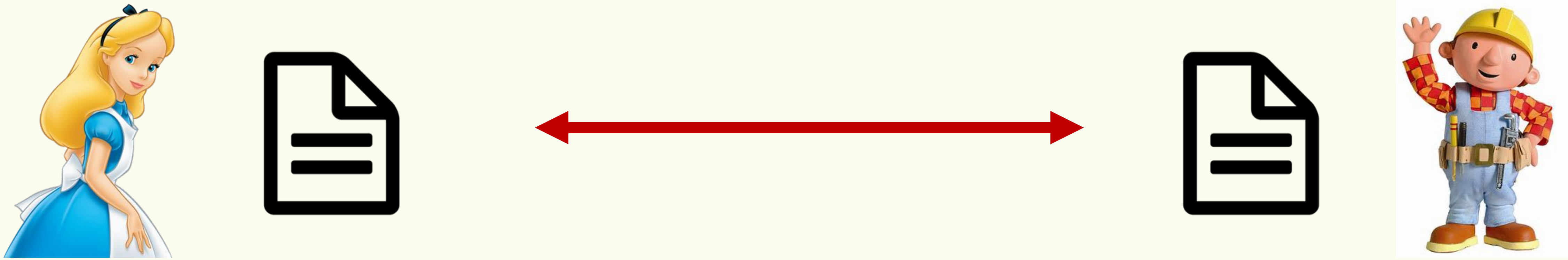
Theorem. (Schwartz-Zippel) A **non-zero** polynomial $a(X)$ mod p of degree d has at most d zeros.

If we pick x uniformly at random from $\mathbb{Z}_p$ and $a(X)$ has degree d, what what can we ay about $\mathbb{P}(a(x) = 0)$?

$$\mathbb{P}(a(x) = 0) \leq \frac{d}{p}$$

File Comparison

1 GB file A 1 GB file B



File Comparison Protocol

- Alice and Bob agree on a prime p
- Alice encodes A as a sequence $(a_0, a_1, \ldots, a_d)$ of elements of $\mathbb{Z}_p$
  - Let $a(X) = a_0 + a_1 X + \cdots + a_d X^d$
- Bob encodes B as a sequence $(b_0, b_1, \ldots, b_d)$ of elements of $\mathbb{Z}_p$
  - Let $b(X) = b_0 + b_1 X + \cdots + b_d X^d$
- Alice picks a random $x \in \mathbb{Z}_p$ and sends $a^* = a(x)$ and x to Bob
- Bob checks whether $a^* = b(x)$
  - *If so, Bob says "equal"*
  - *If not, Bob says "not equal"*

File Comparison Protocol - Analysis

- Alice encodes A as a sequence $(a_0, a_1, \dots, a_d)$ of elements of $\mathbb{Z}_p$
  - Let $a(X) = a_0 + a_1 X + \cdots + a_d X^d$
- Bob encodes B as a sequence $(b_0, b_1, \dots, b_d)$ of elements of $\mathbb{Z}_p$
  - Let $b(X) = b_0 + b_1 X + \cdots + b_d X^d$
- Alice picks a random $x \in \mathbb{Z}_p$ and sends $a^* = a(x)$ and x to Bob
- Bob checks whether $a^* = b(x)$

If $A = B$

- … then $a(X) = b(X)$
- … then $a^* = a(x) = b(x)$

File Comparison Protocol - Analysis

- Alice encodes A as a sequence $(a_0, a_1, \dots, a_d)$ of elements of $\mathbb{Z}_p$
  - Let $a(X) = a_0 + a_1X + \cdots + a_dX^d$
- Bob encodes B as a sequence $(b_0, b_1, \dots, b_d)$ of elements of $\mathbb{Z}_p$
  - Let $b(X) = b_0 + b_1X + \cdots + b_dX^d$
- Alice picks a random $x \in \mathbb{Z}_p$ and sends $a^* = a(x)$ and x to Bob
- Bob checks whether $a^* = b(x)$

If $A \neq B$

- … then $a(X) \neq b(X)$
- … then $c(X) = a(X) - b(X)$
  non-zero and degree at most d

$$\mathbb{P}(a(x) = b(x)) = \mathbb{P}(c(x) = 0) \leq \frac{d}{p}$$

Example – Parameters

1GB = 2^{30} bytes = 2^{33} **bits**, i.e., there are $2^{2^{33}}$ possible files

- Pick p slightly larger than $2^{32} = 2^{2^5}$
- Then, we can use $d = 2^{28}$
  - Now we have $p^d > 2^{2^{33}}$ possible file to encode. (Is enough!)
- Probability that two files are misidentified as identical
  - At most $\frac{d}{p} \leq \frac{2^{28}}{2^{32}} = 2^{-4} = \frac{1}{16}$
- Alice transmits two integers in $\mathbb{Z}_p$
  - Each takes roughly 32 bits = 4 bytes

More efficiently

- Working with primes is a bit tricky
- Polynomials can e.g., be defined also over appropriate mathematical structure (an “extension field”) where the coefficients are chunks of 8 bytes.
  - Such polynomials can be evaluated super-efficiently
  - Hardware support in modern CPUs.
  - Your phone, your laptop, etc is evaluating such polynomials continuously [Main application: Cryptographic integrity protection of data]

End Class Summary

Here
we are …. .

CSE 312 – Exam

Will cover everything from class, including:

- HW 1-8
- Sections
- Applications: Not quite, but … (see next slide)
  – Not this last week, and a few more things

Applications

- Pairwise-independent hashing
- Naïve Bayes and basic machine learning
- Data compression
- Differential privacy
- Randomized algorithms

Strictly speaking <u>not covered</u> by final, but help practicing materials from class.

Learning tips

- Focus on first principles
  - Especially for discrete probability, what are we really trying to solve?!
  - What is the underlying $(\Omega, \mathbb{P})$? Helps even when you are not asked explicitly to do. It all boils down to this.
- What can I use, what can I not use?
  - Is independence assumed? Do I need to prove it first?
  - If you use a fact / theorem, always think (and state) why the theorem can be used.
  - Make sure never to leave anything unspecified. For example, if you describe multiple random variables, you have to explicitly say how they are corelated with each other.
- What result would you expect? Is what you get meaningful? In the right range?

Learning tips (cont'd)

- There are tons of resources.
- Textbook covers most, but not all of what we have done.
- Ask if unsure about your own resource for practice.