PRIMALITY

RANDOM vs FIXED INPUT

(1) PRIMALITY OF RANDOM INPUTS

In RSA we must generate large random primes (p, q : $|\langle p \rangle| = |\langle q \rangle| \simeq 1024$ or larger).

How many primes are there?

PRIME NUMBER THEOREM (no proof)

Let $\pi(n) = \left| \{ 1 \le p \le n : p \text{ prime} \} \right|$ (# primes in $\{1, \dots, n\}$)

Then $\lim_{n \to +\infty} \pi(n) / (n / \ln n) = 1$

That is, $\pi(n) = \frac{n}{\ln n} + o\left(\frac{n}{\ln n}\right)$

In fact, it can be proved that for $n \ge 17$

$$\pi(n) \ge \frac{n}{\ln n}$$

Therefore, for $n \ge 17$:

$$\Pr(x = \text{RANDOM}\{1, \dots, n\} \text{ is prime}) \ge \frac{1}{\ln n}$$

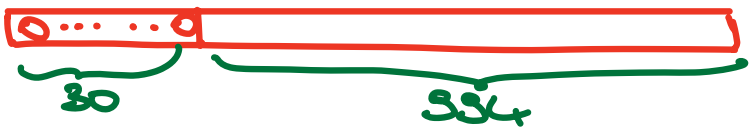
The PNT states that primes are quite dense: there is a prime approximately every $\ln n$ numbers:

$n = 2^{1024}$ (extraction of numbers of 1024 bits)

On average, $\ln(n) = \ln(2^{1024}) \simeq 710$ extractions to obtain a prime.

MORALE: If I have a good primality test for random numbers, I can obtain a large prime $\leq n$ by applying the test over and over to RANDOM$(1, 2, \ldots, n)$

OBSERVATION: We do not have to worry that we extract a prime $p \ll n$! e.g. the possibility that $p < \frac{n}{10^9}$ (thus, represented on $\log n - 30$ bits) is $< 10^{-9}$ (all 30 most significant bits set to 0)



A simple test for checking primality of random numbers

Recall that Fermat's little theorem states that if n is prime:

$$\forall a \in \mathbb{Z}_n^* (= \mathbb{Z}_n^+): \quad a^{n-1} \bmod n = 1$$

This is only a necessary condition! There are composite numbers satisfying:

$$\forall a \in \mathbb{Z}_n^* \ (\neq \mathbb{Z}_n^+): \quad a^{n-1} \bmod n = 1$$

These are called Carmichael's numbers and "behave" like primes on $\mathbb{Z}_n^*$.

Fortunately these numbers are rather rare therefore unlikely to be selected at random.

DEF Given $n \in \mathbb{N}^+$ composite and $a \in \mathbb{Z}_n^*$, we say that n is a base-a pseudoprime if $a^{n-1} \bmod n = 1$

In other words, n "behaves like a prime" w.r.t. a.

OBSERVATION: If $b \in \mathbb{Z}_n^+ - \mathbb{Z}_n^*$ we surely have $b^{n-1} \bmod n \neq 1$ or otherwise

$b \cdot b^{n-2} = b^{n-1} \equiv 1 \bmod n \Rightarrow b^{n-2} = b^{-1}$

When n is composite:

- $a \in \mathbb{Z}_n^*$: $a^{n-1} \equiv 1 \bmod n$ is a "cheater" (makes n "look" like a prime)
- $b \in \mathbb{Z}_n^+$: $b^{n-1} \not\equiv 1 \bmod n$ is a certificate of nonprimality ($b \in \mathbb{Z}_n^*$ or $b \in \mathbb{Z}_n^+ - \mathbb{Z}_n^*$)

For some numbers n, values b: $b^{n-1} \not\equiv 1 \bmod n$ may be difficult to find.

e.g. for Carmichael numbers: $b^{n-1} \not\equiv 1 \bmod n$ only for $b \in \mathbb{Z}_n^+ - \mathbb{Z}_n^*$ (very few)

However, the vast majority of composite numbers admits 2 as a certificate of nonprimality!

POMERANCE THEOREM (1981)

Let X be a random number in $\{1, \ldots, n\}$

Then

$$\text{Prob}(X \text{ is base-2 pseudoprime}) \leq c\, e^{-\frac{1}{2} \ln n \frac{\ln\ln\ln n}{\ln\ln n}}$$

X is base-2 pseudoprime $\iff 2^{X-1} \equiv 1 \bmod X$

$\to 0$ as $n \to \infty$

Based on Pomerance Theorem we can devise this simple test:

```
PSEUDOPRIME(x)
  if MOD_EXP(2, x-1, x) ≠ 1
     then return COMPOSITE
     else return PRIME
```

PSEUDOPRIME is always correct when it returns COMPOSITE but may be wrong when it returns PRIME if X is a base-2 pseudoprime. (ONE-SIDED ERROR)

However, by Pomerance Theorem, if X is randomly chosen in $\{1, \ldots, n\}$:

$$\Pr(\text{PSEUDOPRIME}(x) \text{ is incorrect on random } X \in \{1 \ldots n\}) \leq c\, e^{-\frac{\ln n \ln\ln\ln n}{2 \ln\ln n}}$$

We can use PSEUDOPRIME(x) for the generation of random primes as follows:

```
RANDOM_PRIME(n, s)
  repeat s·ln n times
    x ← RANDOM({1, ..., n^4})
    if PSEUDOPRIME(x) = PRIME
      then return x
  return FAILURE
```

event A ⇒ event B ⇒ $Pr(A) \le Pr(B)$

CORRECTNESS $(n \ge 17)$

1) $Pr(RP(n,s) = \text{FAILURE}) \le$

$\le Pr(\text{extract } s \cdot \ln n \text{ composite numbers}) \le$

$\le \left(1 - \frac{\pi(n)}{n}\right)^{s \ln n} \le \left(1 - \frac{n}{\ln n \cdot n}\right)^{s \ln n}$ ($\pi(n) \ge \frac{n}{\ln n}$)

$= \left(1 - \frac{1}{\ln n}\right)^{s \ln n} \le e^{-s}$ ← $\left(1 - \frac{1}{t}\right)^t \le e^{-1}$

We can make his probability very small by choosing large enough s. E.g., $s = 21 \Rightarrow < 10^{-9}$

2) $Pr(RP(n,s) \text{ is composite}) =$

$= Pr(RP(n) \text{ returns a base-2 pseudoprime})$

$\le Pr(\bigcup_{i=1}^{s \ln n}$ "at iteration i RANDOM$(\{1,\dots,n^4\})$ is a base-2 pseudoprime") $= Pr(\bigcup_{i=1}^{s \ln n} E_i)$

$\le s \cdot \ln n \cdot Pr(\text{RANDOM}(\{1,\dots,n^4\}) \text{ is a base-2 pseudop.})$

↑ union of $s \cdot \ln n$ events

$\leq s \ln n \, e^{-\frac{1}{2}\frac{\ln n \ln\ln n}{\ln\ln n}} \to 0$ for $n \to \infty$

(for 1024-bit numbers, $s = 21$:

$\Pr(RP(n) \text{ is composite}) < 10^{-43}$!)

Clearly

$\Pr(RP(n,s) \text{ is incorrect}) \leq 1 + 2$

($< 10^{-9} + 10^{-43}$ for 1024-bit numbers and $s = 21$)

RUNNING TIME

Each iteration of $RP(n)$ executes MOD_EXP$(z, x-1, x)$, $x \leq n$, running in time $O(|\langle n \rangle|^3)$. Since $s \ln n = O(s|\langle n \rangle|)$:

$$T_{RP}(|\langle n \rangle|, s) = O(s|\langle n \rangle|^4)$$

Worst-case analysis very conservative. In practice much faster (use speed-up hacks, e.g. test for divisibility by small factors)

MILLER-RABIN'S (MR) PRIMALITY TEST

The test will check whether a specific number $N > 1$ is prime or composite.

OBSERVATION: Very different from the PSEUDOPRIMALITY test for random inputs that uses a deterministic algorithm:

- always correct on a large fraction of the inputs
- always wrong on a small set of inputs (base-2 pseudoprimes)

The MR primality test works in the worst case (e.g., even if the input is a base-2 pseudoprime or a Carmichael number).

The MR test is a randomized algorithm: it makes use of RANDOM(·) to check for primality/compositeness.
The correctness analysis is performed w.r.t the probability space induced by the calls to RANDOM(·)

MAIN IDEA : Given a fixed $n > 1$:

- randomized search for a "certificate" of nonprimality (compositeness).

- The search for such a certificate is repeated s times. The probability of "missing" such a certificate when n is composite is $\leq \frac{1}{2^s}$ (becomes quickly very small)

- If no certificate is found in the s iterations, MR(n) declares that n is prime

- MR is a randomized, one-sided error algorithm :
  1. always correct if n is prime
  2. may be incorrect if n is composite (but with probability that can be made very small)

CERTIFICATES OF NONPRIMALITY USED IN MR(n)

1. We look for random values $a \in \mathbb{Z}_n^+$:
   $a^{n-1} \not\equiv 1 \bmod n$
   (in other words, look for a : n is not a base-a pseudoprime)

REMARK: This extends the test for random numbers based on base-2 pseudoprimes to arbitrary values of $a \in \mathbb{Z}_n^+$. Clearly this is not sufficient for "hard" inputs:

Carmichael's numbers n

$\forall a \in \mathbb{Z}_n^*: \; a^{n-1} \equiv 1 \bmod n$.

For these numbers, the only valid certificates may come from

$a \in \mathbb{Z}_n^+ - \mathbb{Z}_n^*$ (since $a^{n-1} \not\equiv 1 \bmod n$)

but these values may be very few w.r.t n (in fact, Carmichael numbers tend to have very few prime factors $\Rightarrow$ $|\mathbb{Z}_n^+ - \mathbb{Z}_n^*| \ll n$)

2. Look for nontrivial square roots of unity:

$a \in \mathbb{Z}_n^+ : (a \not\equiv \pm 1 \bmod n) \wedge (a^2 \equiv 1) \bmod n$

(these are numbers $a \in \mathbb{Z}_n^* : a = a^{-1} \bmod n$)

Why?

THEOREM:

If $\exists a \in \mathbb{Z}_n : (a \not\equiv \pm 1 \bmod n) \wedge (a^2 \equiv 1) \bmod n$

$\Rightarrow$ n is composite

EXAMPLE: 4 in $\mathbb{Z}_{15}$: $(4 \not\equiv \pm 1 \bmod 15) \wedge (4^2 = 16 \equiv 1 \bmod 15)$

PROOF: We prove the counterpositive:

n prime $\Rightarrow$ the only values $d \in \mathbb{Z}_n$:
$d^2 \equiv 1 \bmod n$ are $d \equiv \pm 1 \bmod n$

Consider the congruent equation:

$$x^2 \equiv 1 \bmod n \iff$$
$$x^2 - 1 \equiv 0 \bmod n \iff$$
$$(x+1)(x-1) \equiv 0 \bmod n \iff$$
$$(x+1)(x-1) = Kn, \quad K \in \mathbb{N}$$

Thus

$$n \mid (x+1)(x-1).$$

Since n is prime:

$n \mid ab \Rightarrow (n \mid a) \vee (n \mid b)$

or otherwise $\gcd(n, a) = 1$ and $\gcd(n, b) = 1$ therefore $\gcd(n, ab) = 1$ (Bezout) which contradicts $n \mid ab$ ($\Rightarrow \gcd(n, ab) = n$)

Therefore

$n \mid (x+1)$ or $n \mid (x-1)$

$n \mid (x+1) \Rightarrow x+1 = K_1 n \Rightarrow x = -1 + K_1 n$

$n \mid (x-1) \Rightarrow x-1 = K_2 n \Rightarrow x = 1 + K_2 n$

It must be

$$x \equiv \pm 1 \bmod n$$

Q.E.D.