HAI709I : Fondements cryptographiques de la sécurité, Université de Montpellier, 2024

02/12/2024. Lecture 11.

1 Density of the prime numbers

A natural number $p \in \mathbb{N}$ is called prime if it has exactly two different natural divisors: 1 and p. The list of prime numbers begins with

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31, 37,\ 41,\ 43,\ 47, 53,\ 59, \ldots$$

There are quite *many* prime numbers. This statement can be mode more precise in different ways:

- the set of prime numbers is infinite (this theorem was known to Euclid)
- moreover, for every integer number $n > 0$, there exists a prime number p such that $n \le p < 2n$ (this property is called Bertrand's postulate; it was proven by Chebyshev)

Denote $\pi(n)$ the prime-counting function (the number of primes less than or equal to n). Then

- there exist numbers $c_1 > 0$ and $c_2 > 0$ such that for all n
$$c_1 \cdot \frac{n}{\ln n} < \pi(n) < c_2 \cdot \frac{n}{\ln n}$$
(so-called Chebyshev's bounds)
- and moreover, for every $\epsilon > 0$ there exists an $n_0 = n_0(\epsilon)$ such that for all $n > n_0$
$$(1-\epsilon)\frac{n}{\ln n} < \pi(n) < (1+\epsilon)\frac{n}{\ln n}$$
(proven by Hadamard and de la Vallée Poussin).

In the class we used the bound proven by Hadamard and de la Vallée Poussin to deduce the following property:

Proposition 1. *There exist a $c > 0$ and a $k_0 > 0$ such that for all integer numbers $k > k_0$ the number of primes between 2^k and 2^{k+1} is greater or equal to $c \cdot 2^k/k$.*

This proposition means that if we choose at random an integer number x whose binary expansion consists of 1 followed by k binary digits,

$$x = 2^k \cdot 1 + 2^{k-1}b_{k-1} + \ldots + 4b_2 + 2b_1 + b_0, \text{ where } b_i \in \{0,1\} \text{ for each } i$$

(a number between 2^k and 2^{k+1}), then it will be prime with a probability of at least

$$\frac{c \cdot 2^k/k}{2^k} \ge c/k.$$

This observation shows how we can produce large prime numbers: we pick up random integer numbers and test their primality, until we find a number that is actually prime. Indeed, if we sample one natural number between 2^{k-1} and 2^k, it will be prime with a probability $\ge c/k$. This probability is small but not negligibly small. So we can improve the probability of success if we repeat this experience many times. If we take at random **const** $\cdot\, k$ integer numbers with k binary digits (for a large enough factor **const**), then *at least one* of these numbers will be prime with a probability of > 0.99. What remains missing in this scheme is an efficient test of primality. We will discuss such a test in the next section.

2 The Miller–Rabin primality test

For an integer number n we can verify whether it is prime by trying all potential divisors among the candidates $1, 2, \ldots, \lfloor\sqrt{n}\rfloor$. However, this procedure is very slow. If the binary representation of n consists of k digits (i.e., $2^{k-1} \leq n < 2^k$), then the number of candidates $\sqrt{n} \sim \sqrt{2^k} = 2^{k/2}$ is exponential in k. In what follows we discuss a much more efficient (polynomial time) test of primality.

Let us recall Fermat's little theorem.

Theorem 1. *For all prime numbers p and for all $a \in \{1, 2 \ldots, p-1\}$*

$$a^{p-1} = 1 \mod p.$$

Proof. Let us multiply a by each number from the list $1, 2, \ldots, p-1$,

$$\begin{array}{rcl} a \cdot 1 & = & b_1 \mod p \\ a \cdot 2 & = & b_2 \mod p \\ \ldots & & \\ a \cdot (p-1) & = & b_{p-1} \mod p \end{array}$$

We know that for $i \neq j \mod p$ we have $a \cdot i \neq a \cdot j \mod p$. Therefore, all numbers b_i are pairwise different. Hence, in the list

$$b_1, b_2, \ldots, b_{p-1} \mod p$$

every number $1, 2, \ldots, p-1$ appears exactly once. In other words, $b_1, b_2, \ldots, b_{p-1}$ is a permutation of the list of numbers $1, 2, \ldots, (p-1)$. It follows that in the product

$$(a \cdot 1) \cdot (a \cdot 2) \cdot \ldots \cdot (a \cdot (p-1)) = b_1 \cdot b_2 \cdot \ldots \cdot b_{p-1} \mod p$$

the factors $1, \ldots, p-1$ appear once in left-hand side and in the right-hans side. Simplifying the equality we obtain

$$\underbrace{a \cdot \ldots \cdot a}_{p-1} = 1 \mod p,$$

and the theorem is proven. □

This observation motivates the following naive test of primality:

Fermat test `of primality for an integer number` n :

1. take a random number $a \in \{1, 2 \ldots, n-1\}$
2. compute $b \leftarrow (a^{n-1} \mod n)$
3. if $b = 1 \mod n - 1$, return "prime" ; otherwise return "not prime".

Fermat's little theorem implies that for all prime numbers the Fermat test always gives the right answer and says *prime*. Is it true that for every composite number the test says *not prime* with a non-negligible probability? Unfortunately, this is not always the case. There exist composite integer numbers (cf. *Carmichael numbers*) for which this test fails for all $a \in \{1, 2 \ldots, n-1\}$.

Fortunately, there are other tests that work correctly for all number, and in what follows we discuss one of these tests. This test is randomized. For each prime number it always returns the right answer (with probability one), and for each composed number it returns the right answer with a probability at least $1/2$.

Miller–Rabin test `of primality for an integer number` $n > 1$ **:**

1. denote $n - 1 = m \cdot 2^r$, where m is an odd number (the maximal odd divisor of $n - 1$)
2. take a random number $a \in \{1, 2 \ldots, n - 1\}$
3. compute the series of numbers
   - $b_0 := a^m \mod n$
   - $b_1 := (b_0^2) = a^{m \cdot 2} \mod n$
   - $b_2 := (b_1^2) = a^{m \cdot 4} \mod n$

     $\vdots$
   - $b_r := (b_{r-1}^2) = a^{m \cdot 2^r} = a^{n-1} \mod n$

   (we assume that each b_i belongs to $\{0, 1, 2, \ldots, n - 1\}$)
4. if $b_0 = b_1 = \ldots = b_r = 1$, then return "prime"
5. if there exists an $i \in \{0, 1, \ldots, r - 1\}$ such that $b_i = p - 1$ (equivalently, $b_i = -1 \mod n$), return "prime"
6. in all other cases return "not prime"

In the class we proved the following statements:

- if n is prime, the Miller–Rabin test says "prime" with probability 1
- if n has at least at least two different prime factors, then the Miller–Rabin test says not "prime" with probability $\geq 1/2$.

We did not proved in the class the fact that the probability of failure is small for n that are powers of prime numbers (such a number is not prime but it has only one prime factor). However, the property "n is a power of an integer number" can be tested deterministically in polynomial time.

Theorem 2. *If n is a prime number, then the Miller–Rabin test returns "prime" with probability* 1.

Sketch of the proof. First of all, from Fermat's little theorem it follows that for every a

$$a^{n-1} = a^{m \cdot 2^r} = 1 \mod n.$$

Thus, $b_r = 1 \mod n$, and the list of the values $(b_0, b_1, \ldots, b_r)$ can look like

$$(1, 1, 1, \ldots, 1)$$

or

$$(*, * \ldots, *, -1, 1, \ldots, 1)$$

or

$$(*, * \ldots, *, 1, 1, \ldots, 1)$$

where $*$ denotes any number that is not equal to $\pm 1 \mod n$. In the first and the second case, the test returns the answer "prime". It remains to show that the third case is impossible.

The third case above means that for some i we have $b_i \neq \pm 1 \mod n$, and $b_{i+1} = 1 \mod n$. Combining this with the fact $b_{i+1} = b_i^2 \mod n$, we see that the equation

$$x^2 = 1 \mod n$$

has at least three different roots: 1, -1, and b_i. However, modulo a prime number n, every polynomial of degree 2 cannot have more than 2 roots. We have arrived to a contradiction, which completes the proof. □

Theorem 3. *If n has at least two different prime factors, then the Miller–Rabin test returns "not prime" with a probability $\geq 1/2$.*

Sketch of the proof. If n has at least two different prime factors, than it can be represented as a product $n = n' \cdot n''$ where n' and n'' are co-prime integer numbers strictly greater than 1.

Let i_0 denote the maximal integer number such that there exists at least one $a \in \{1, \dots, n-1\}$ such that

$$a^{m \cdot 2^{i_0}} = -1 \mod n.$$

(Observe that such an i_0 exists: we know for sure that $(p-1)^m = (-1)^m \mod n = -1 \mod n$ since m is odd; thus, $-1 \mod p$ can appear even in the very first position of the list $(b_0, b_1, \dots, b_r)$). Further, let us set

$$H := \{a \in \{1, 2, \dots, n-1\} \text{ such that } a^{m \cdot 2^{i_0}} = \pm 1 \mod n\}.$$

Observe that the Miller–Rabin test can return the (false) answer "prime" for the input n *only* if the randomly chosen a belongs to H; if $a \notin H$, then the test will return the correct answer. Therefore, to show that the probability of an error is $\leq 1/2$, we need to prove that the set H covers at most a half of all integers $\{1, 2, \dots, n-1\}$.

Claim 1: *there exists at least one a that does not belong to H.* Indeed, let us fix an a such that

$$a^{m \cdot 2^{i_0}} = -1 \mod n.$$

Observe that $a^{m \cdot 2^{i_0}} = -1 \mod n'$. Now we inspect the list of numbers

$$a,\ a + n',\ a + 2n',\ a + 3n',\ \dots, a + (n'' - 1)n'$$

We make two simple observations.

Fact 1. For each number $\tilde{a}$ in this list we have $\tilde{a}^{m \cdot 2^{i_0}} = -1 \mod n'$ (since all these numbers are equal to each other modulo n').

Fact 2. All these numbers are pairwise distinct modulo n'' (since the difference between every two numbers $a + in'$ and $a + jn'$ is equal to $(i - j)n'$, which is not divisible by n'').

From Fact 2 it follows that the list contains every possible reminder modulo n'' exactly once, so there must be an element $a + jn'$ that is equal to 1 modulo n''. We take this number as $\hat{a}$. By the construction, we have

$$\hat{a} = -1 \mod n' \text{ and } \hat{a} = 1 \mod n''.$$

It is clear that $\hat{a} \neq \pm 1 \mod n$. So, we have found and $\hat{a} \notin H$.

Claim 2: *the set H covers at most a half of the set* $\{1, 2, \ldots, p-1\}$. To prove this claim we denote the elements of H as follows:

$$H = \{h_1, \ldots, h_s\}.$$

Let us multiply each of these numbers by $\hat{a}$ chosen above:

$$\hat{a} \cdot h_1 \mod n, \ldots, \hat{a} \cdot h_s \mod n$$

It is not hard to see that all these numbers are pairwise different module n (the difference between them is not divisible by n), and each of them is not in H. Hence, in the complement set $\{1, 2, \ldots, n-1\} \setminus H$ we have at least as many elements as in H. Therefore, when we choose at random an element in $\{1, 2, \ldots, n-1\}$, with a probability $\geq 1/2$, we get an element *not* in H, and the test returns the answer *not prime*. □

Exercise 1. Construct a polynomial time deterministic algorithm that takes as input a binary representation of a number n and tests whether n is a power of an integer, i.e., whether there exist integer numbers m and $k > 1$ such that $n = m^k$. If such m, k exist, the algorithm should find them.

Remark 1. The Miller–Rabin test uses randomness. Can we test primality of integer numbers deterministically, without any probability to get a wrong answer? The answer to this question is *yes*. The algorithm invented by Agrawal, Kayal, and Saxena is deterministic, and it verifies primality of a given integer number in polynomial time. However, in practice the algorithm by Agrawal–Kayal–Saxena is slower than the Miller–Rabin test.