Advanced Algorithms	Lecture 2 : March 8, 1995
Lecturer: Johan Håstad	*Scribe: Mats Näslund*

2.1 Primality testing cont'd.

2.1.1 The Miller-Rabin primality test cont'd.

We implement the Miller-Rabin test as follows. Assume $N \sim 2^n$.

Miller-Rabin$(N, k) =$
  find s such that $N - 1 = 2^s t$, t odd
  repeat k times
    choose $a \in_U \mathbb{Z}_N^*$
    define the sequence $\{u_i\}_{i=0}^{s}$ by:
    $u_0 := a^t \pmod N$
    $u_i := u_{i-1}^2 \pmod N$, $i := 1, 2, \ldots, s$
    if $\exists\, i : 1 \leq i \leq s$ s.t. $u_i = 1$ and $u_{i-1} \neq \pm 1$ **then**
     return "not prime"
  return "probably prime"

Claim 1 *If N is prime, the algorithm decides correctly. If N is composite, the probability that the algorithm is incorrect is bounded above by 2^{-k}.*

The first part is obvious. If N is composite, we will show that for $k = 1$ the probability of returning "probably prime" is at most $1/2$. Since the k a's are chosen randomly and independently, the claim follows.

Before proving the second part, let us note that if we for instance choose $k = 100$, the algorithm answers correctly with an overwhelming probability: $1 - 2^{-100}$. Also observe that the running time is $O(kn^3)$, since for each of the k a's, we compute each of the at most n u_i's by a simple squaring.

The idea behind the proof of claim 1 is to distinguish two special cases:
1. N is a prime power, $N = p^b$, $b \geq 2$ and

2. N has at least two distinct prime factors.

Proof: (In case 1.) Here, we are almost home free, since this case won't pass the criterion in Fermat's little theorem, i.e. for most a we have $a^{N-1} \not\equiv 1 \pmod{N}$. Assume $N = p^b$, $b \geq 2$. If indeed, $a^{N-1} \equiv 1 \pmod{N}$, we find many a_i such that $a_i^{N-1} \not\equiv 1 \pmod{N}$. This can be done by noting that if $a^{N-1} \equiv 1 \pmod{N}$ and we set $a_i = (a + ip)$ where $1 \leq i < p$, then:

$$\begin{aligned}
a_i^{N-1} &\equiv (a+ip)^{N-1} \equiv \sum_{j=0}^{N-1} \binom{N-1}{j} a^{N-1-j}(ip)^j \\
&\equiv a^{N-1} + (N-1)ipa^{N-2} + p^2 \sum_{j=2}^{N-1} \binom{N-1}{j} a^j (ip)^{j-2} \\
&\equiv a^{N-1} + (N-1)ipa^{N-2} \\
&\equiv 1 + (N-1)ipa^{N-2} \pmod{p^2}.
\end{aligned}$$

But $(N-1)ipa^{N-2} \not\equiv 0 \pmod{p^2}$ since $p \nmid N-1$, $p \nmid i$ and $p \nmid a^{N-2}$. Remark: It can in fact be shown that if $N = p^b$, then

$$\Pr_{a \in \mathbb{Z}_N}[a^{N-1} \equiv 1 \pmod{N}] \approx \frac{1}{p^{b-1}}.$$

■

To prove claim 1 in case two, i.e. when N is not a prime power, we need the following three lemmas.

Lemma 1 *Let p be a prime where $p - 1 = 2^{r'}t'$ where t' is odd. Then*

$$\Pr_{a \in \mathbb{Z}_p^*}[a^{2^r t} \equiv -1 \pmod{p} \mid a^{2^s t} \equiv 1 \pmod{p}]$$

where t is odd is bounded by $\max(2^{r-r'}, 2^{r-s})$ *when* $r < \min(r', s)$ *and 0 otherwise.*

Lemma 2 *Let p be a prime where $p - 1 = 2^{r'}t'$ where t' is odd. Suppose $s < r'$ then*

$$\Pr_{a \in \mathbb{Z}_p^*}[a^{2^s t} \equiv 1 \pmod{p}]$$

where t is odd is bounded by $2^{s-r'}$.

Lemma 3 *Let p be a prime where $p-1 = 2^{r'}t'$ where t' is odd. Then*

$$\Pr_{a \in \mathbb{Z}_p^*}[a^t \equiv 1 \pmod{p} \mid a^{2^s t} \equiv 1 \pmod{p}]$$

where t is odd is bounded by $\max(2^{-r'}, 2^{-s})$.

Proof: (Of lemma 1, 2 and 3.) We leave this as an exercise. As a hint, you might consider the fact that the multiplicative group mod p is isomorphic to the additive group mod $p-1$. In this isomorphism 1 corresponds to 0 and -1 to $(p-1)/2$. Squaring in the multiplicative setting corresponds to multiplication by 2 in the additive setting and you can then analyze what power of two divides the numbers in question.

The following intuition (which can be proved) might be of help: If $x^{2l} \equiv 1 \pmod{p}$, what was $x^l \pmod{p}$? The only roots of $1 \pmod{p}$ are ± 1. It seems natural that half of the roots are 1 and half -1 and this is also the actual case! ■

We now proceed with the proof of claim 1.

Proof: (In case 2.) Here assume $N = p_1 p_2 M$ for some $M \in \mathbb{Z}$. As before, let $N-1 = 2^s t$, t odd. For N to pass the test as "probably prime", we must have $u_0 \equiv 1 \pmod{N}$ *or* $u_r \equiv -1 \pmod{N}$ for some $r < s$. These events are mutually exclusive.

(i) If $u_0 \equiv 1 \pmod{N}$, by the Chineese remainder theorem: $u_0 \equiv 1 \pmod{p_1}$ *and* $u_0 \equiv 1 \pmod{p_2}$. Hence $u_s \equiv 1 \pmod{p_1}$ and $u_s \equiv 1 \pmod{p_2}$ too.

(ii) If $u_r \equiv -1 \pmod{N}$ for some $r < s$, by the same reasoning, $u_r \equiv -1 \pmod{p_1}$ and $u_r \equiv -1 \pmod{p_2}$ which also implies $u_s \equiv 1 \pmod{p_1}$, $u_s \equiv 1 \pmod{p_2}$.

The error probability is at most the sum of the probabilities of the two events (i) and (ii) above. Let us analyze (ii) first.

If we chose a random $a \pmod{N}$, then we also choose $a \pmod{p_1}$, $a \pmod{p_2}$ uniformly and independently by previous arguments (lecture 1). Assume $p_1 - 1 = 2^{r_1} t_1$ and $p_2 - 1 = 2^{r_2} t_2$ with t_1, t_2 odd. Without loss of generality, assume $r_1 \geq r_2$. We distinguish two separate cases:

(a) $s < r_1$. Then by lemma 2

$$\Pr_{a \in \mathbb{Z}_{p_1}^*}[a^{2^s t} \equiv 1 \pmod{p_1}] \leq 2^{s - r_1} \leq \frac{1}{2} \qquad (*)$$

and we are clearly done since this also covers case (i).

(b) $s \geq r_1$. Suppose $a^{2^s t} \equiv 1$ both mod p_1 *and* mod p_2. Then, by lemma 1, the probability of having $a^{2^r t} \equiv -1 \pmod{p_1}$ *and* $a^{2^r t} \equiv -1 \pmod{p_2}$ is, by the independence, bounded by $2^{r-r_1}2^{r-r_2} = 2^{2r-r_1-r_2}$ if $r < r_2$ and 0 otherwise. Thus the probability of having -1 simultaneously for *some* r is at most

$$\sum_{r=0}^{r_2-1} 2^{2r-r_1-r_2} \leq 2 \cdot \underbrace{2^{-2-(r_1-r_2)}}_{\text{the last term}} . \qquad (**)$$

Finally, let us take into account the other possible event, (i). The probability that

$$u_0 \equiv 1 \pmod{p_1} \qquad \text{and} \qquad u_0 \equiv 1 \pmod{p_2}.$$

is by lemma 3 bounded by $2^{-(r_1+r_2)}$, since $s \geq r_1 \geq r_2$. Adding this to $(**)$ gives a total error probability of at most

$$\frac{1}{2}\left(2^{-(r_1-r_2)} + 2^{1-(r_1+r_2)}\right) < \frac{1}{2}.$$

■

Some notes on the algorithm

If there are lots of a's that can be used as witnesses of the compositeness of N, can we say something about a in advance? Is there some set of a's that *guarantees* revealing the compositness?

There exists a number theoretic conjecture known as the *extended Riemann hypothesis*, ERH. (See [5] for details.)

Claim 2 *Under the ERH, there always exists a witness $a \leq 2(\log N)^2$ of the compositness of N.*

Corollary 1 *Under the ERH, the Miller-Rabin test gives a* deterministic *test of primality that runs in time $O((\log N)^5)$.*

Simply try all $a \leq 2(\log N)^2$.

2.1.2 Other tests for primality

There are several other, more or less practical tests. We list a few by their inventors. Below, n equals $\log_2 N$.

Solovay & Strassen, [9]. Define the *Legendre symbol* $(\frac{b}{p}) = 1$ if the equation $x^2 \equiv b \pmod{p}$, p prime, has a solution, -1 otherwise. The *Jacobi symbol* $(\frac{b}{N})$ when $N = \prod p_i^{e_i}$ is defined as the product of the Legndre symbols $(\frac{b}{p_i})^{e_i}$ for each prime factor p_i in N. The Jacobi symbol can be efficiently computed even if the factorisation of N is unknown.

It can be shown that *if* N is prime then

$$b^{(N-1)/2} \equiv \left(\frac{b}{N}\right) \pmod{N} \tag{$*$}$$

but if N is composite then the probability that $(*)$ holds for a random b is at most $1/2$. This criterion is in fact stronger than the criterion in Fermats little theorem, since there are no analogs of Carmichael numbers with respect to this test.

Adleman, Pomerance & Rumely, [2]. This is a deterministic algorithm that uses Gauss-sums anf runs in time $O(n^{\log\log n})$. The algorithm is not practical.

Cohen & Lenstra, [3]. This algorithm is based on the above and is in fact possible to implement. The algorithm uses Jacobi-sums.

Goldwasser & Killian, [4]. This is a probabilistic algorithm that *always* gives a correct answer. It runs in expected polynomial time. It is provably correct for "most" p. The basic idea is simple. It can be described as: "If a has an order mod N that is a non-trivial divisor in $(N-1)/2$ then: $(N-1)/2$ is prime implies N is prime." As an example, 4 has order 5 mod 11 so if 5 is a prime then 11 is prime too! The crux is that it is very unlikely that $(N-1)/2$ be prime. As an example, the statement "If 8 is prime then 17 is prime" is of no great help. This problem was circumvented by Goldwasser & Killian by working with elliptic curves instead of the integers mod N.

Adleman & Huang, [1]. This is a "fix" of Goldwasser & Killian that makes it work for all N. It uses so called Jacobians of elliptic curves. The idea is slightly counter-intuitive. It reduces the problem of deciding primality for a certain number N to the problem of deciding primality for a number N' that is *larger* than N ($N' \sim N^2$). It would seem that nothing is gained by this, but what happens is that N' will be random. Thus it is possible to run Goldwasser & Killian on this number instead.

2.2 Factoring

So, suppose that we have fed N to some algorithm for primality testing and that we got the answer "N is composite". Let us factor N into primes! What methods are available?

2.2.1 The "naive" algorithm

If N is composite, it must have at least one prime factor p with $p \leq \sqrt{N}$. Try by dividing N by all integers $p \leq \sqrt{N}$. The running time is obviously $O(\sqrt{N}) = O(2^{n/2})$.

2.2.2 Pollard's ρ-method

We start by giving the algorithm first and then discuss it. The algorithm finds one non-trivial factor $d|N$. It is of course easy to proceed and factor N completely.

Pollard(N) =
  let $x_0 \in_U \mathbb{Z}_N$
  define the sequence $\{x_i\}_{i\geq 1}$ by $x_{i+1} \equiv x_i^2 + 1 \pmod{N}$
  for $i := 1, 2, \ldots$ **do**
    $d := \gcd(x_{2i} - x_i, N)$
    if $d > 1$ **then**
     d is a non-trivial factor in N, **stop**

The big questions are: How can we be sure that we *will* find a non-trivial factor and if so, how long do we have to wait?

Assume that $p|N$. If the sequence $\{x_i\}_{i\geq 1}$ behaves "randomly" so that x_{2i}, x_i can be considered as random integers mod N, then

$$\Pr[p \text{ divides } x_{2i} - x_i] = \frac{1}{p}$$

which in turn implies that $p|\gcd(x_{2i} - x_i, N)$.

Claim 3 *A non-trivial factor p will be found after $O(\sqrt{p}) \in O(N^{1/4})$ steps.*

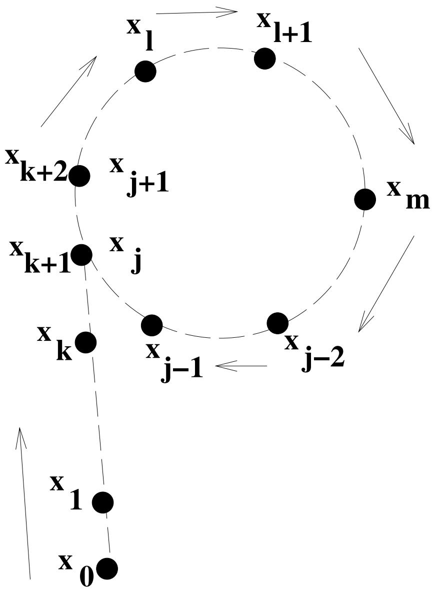

Figure 2.1: The sequence $x_i \pmod p$.

Proof: Let us study the sequence $\{x_i\}_{i\geq 1} \bmod p$. We can do this graphically, fig 2.1[1]: After the j:th step, the sequence repeats! We can look at it as having two runners, x_{2i} and x_i running in a circle, one with twice the speed of the other. Eventually x_{2i} will be passing x_i. It is when this passing occurs that the factor d falls out. Since $x_{2i} \equiv x_i \pmod p$, $p | x_{2i} - x_i$.

What remains is to estimate the number of steps untill this occurs, the "circumference" of the circle. This follows from the lemma below. ■

Lemma 4 *If $f : \mathbb{Z}_p \mapsto \mathbb{Z}_p$ behaves like a random function and we define the sequence $\{x_i\}_{i\geq 0}$ by $x_{i+1} = f(x_i)$ with x_0 arbitrary we have:*

$$E_f[\text{ smallest } j \text{ s.t. } x_j \equiv x_i \pmod p,\ i < j] \in O(\sqrt{p}).$$

Proof: Exercise. ■

Now, could we have chosen f in some other way, say by computing $x_{i+1} \equiv \lfloor N \sin(2\pi x_i / N) \rfloor \pmod N$? The answer is no! To be able to apply the reason-

[1]The name of this algorithm should be obvious from this figure.

ing mod p we need an algebraic structure on the sequence. We want to be able to state things like $x_j \equiv x_i \pmod{N} \Rightarrow x_{j+1} \equiv x_{i+1} \pmod{p} \Rightarrow x_{j+1}^2 + 1 \equiv x_{i+1}^2 + 1 \pmod{p}$ and so on. We don't have this structure if the sequence is truly random.

Why the choice of the specific function $x \mapsto x^2 + 1 \pmod{N}$? The reason for this is basically heuristic: It works well in practice. It is also easy to compute. Note that we could not have used functions like $x \mapsto ax + b \pmod{N}$. Such functions are $1-1$. We leave as an exercise to estimate the "return time", i.e. number of steps until the sequence $\{x_i\}$ repeats in this case.

Let us end by an example of an execution of the ρ-method.

Example 1 *Let us factor $N = 1387 = 19 \cdot 37$. Start by $x_0 = 0$. Let $y_i \equiv x_{2i} - x_i \pmod{1387}$.*

i:	1	2	3	4	5	6	7	8	9	10
x_i	1	2	5	26	677	620	202	582	297	829
x_{2i}	2	26	620	582	829	$\cdots$				
y_i	1	24	615	556	152	$\cdots$				
$\gcd(y_i, 1387)$	1	1	1	1	19	$\cdots$				

After step 5, the factor 19 falls out. Observe that $\sqrt{19} \approx 4.36$.

References

[1] Adleman & Huang: *Primality Testing and two-dimensional Abelian Varieties over Finite Fields.* LNCS, vol 1512, Springer-Verlag, 1992.

[2] Adleman, Pomerance & Rumely: *On Distinguishing Prime Numbers from Composite Numbers.* Annals of Math. 117, 1983.

[3] Cohen & Lenstra: *Primality Testing and Jacobi Sums.* Mathematics of Computation 42, 1984.

[4] Goldwasser & Killian: *Almost all Primes can be Quickly Certified.* STOC 1986.

[5] Ireland & Rosen: *A Classical Introduction to Modern Number Theory.* Springer-Verlag 1990.

[6] Miller: *Riemann's Hypothesis and Tests for Primality.* 7th STOC.

[7] Pollard: *A Monte Carlo Method for Factorization.* BIT 15, 1975.

[8] Rabin: *Probabilistic Algorithms for Testing Primality.* J. Number Theory 12, 1980.

[9] Solovay & Strassen: *A Fast Monte Carlo Test for Primality.* SIAM J. on Comp. 6, 1977.