Math 445 Number Theory

September 3, 2004

Miller-Rabin Test: Given a number N, and a base a, compute $N-1 = 2^k \cdot d$, with d odd. Then compute

$a_0 = a^d \pmod N$, $a_1 = a^{2d} = (a^d)^2 \pmod N$, $a_2 = (a_1)^2 \pmod N$, $\dots$, $a_k = a^{2^k d} = a_{k-1}^2 \pmod N$

If $a_0 = 1$ or $a_i \equiv -1 \pmod N$ for some $i \leq k-1$, then N passes the test; it is either prime or a *strong pseudoprime* to the base a. If not, then N is definitely not prime.

Monier and Rabin in 1980 showed that a composite number N is a strong pseudoprime for at most $1/4$ of possible bases a. So if N passes this test for m randomly chosen bases $a_1, \dots, a_m$, then N has only a 1 in 4^m chance of *not* being prime. That is, multiple Miller-Rabin tests are very good at ferreting out non-primes.

If this test tells us that a number N is composite, how do we find its factors? The most straightforward approach; test divide all numbers less than $\sqrt{N}$, or better, all *primes* less than $\sqrt{N}$; eventually you will find a factor. But this requires on the order of $\sqrt{N}$ steps, which is far too large.

A different method uses the fact that if $N = ab$ and $a_1, \dots a_n$ are chosen at random, a is more likely to divide one of the a_i (or rather (for later efficiency), one of the differences $a_i - a_j$), than N is. This can be tested for by computing gcd's, $d = (a_i - a_j, N)$; this number is $1 < d < N$ if a (or some other factor) divides $a_i - a_j$ but N does not, and finds us a proper factor, d, of N. The probability that a divides none of the differences is approximately $1 - 1/a$ for each difference, and so is approximately

$$(1-\frac{1}{a})^{\binom{n}{2}} = ((1-\frac{1}{a})^a)^{\frac{n(n-1)}{2a}} \approx ((1-\frac{1}{a})^a)^{\frac{n^2}{2a}} \approx ((1-\frac{1}{a})^a)^{\frac{n^2}{2a}} \approx (e^{-1})^{\frac{n^2}{2a}} = e^{\frac{-n^2}{2a}}$$

which is small when $n^2 \approx a \leq \sqrt{N}$, i.e., $n \approx N^{1/4}$. The problem with this method, however, is that it requires $n(n-1)/2 \approx \sqrt{N}$ calculations, and so is no better than trial division! We will rectify this by choosing the a_i *pseudo-randomly* (which will also explain the use of differences). This will lead us to the Pollard ρ method for factoring.