LARGE PRIME NUMBERS

This writeup is modeled closely on a writeup by Paul Garrett. See, for example,

`http://www-users.math.umn.edu/~garrett/crypto/overview.pdf`

1. Fast Modular Exponentiation

Given positive integers a, e, and n, the following algorithm quickly computes the reduced power $a^e \bmod n$. (Here $x \bmod n$ denotes the element of $\{0, \cdots, n-1\}$ that is congruent to x modulo n. Note that this usage of $x \bmod n$ does not denote an element of $\mathbb{Z}/n\mathbb{Z}$ because such elements are cosets rather than coset representatives.)

- (*Initialize*) Set $(x, y, f) = (1, a, e)$.
- (*Loop*) While $f > 0$, do as follows:
  - if $f \bmod 2 = 0$ then replace (x, y, f) by $(x, y^2 \bmod n, f/2)$,
  - otherwise replace (x, y, f) by $(xy \bmod n, y, f-1)$.
- (*Terminate*) Return x.

This algorithm is strikingly efficient both in speed and in space. Especially, the operations on f (halving it when it is even, decrementing it when it is odd) are very simple in binary. To see that the algorithm works, represent the exponent e in binary, say

$$e = 2^g + 2^h + 2^k, \quad 0 \le g < h < k.$$

The algorithm initializes

$$(1,\ a,\ 2^g + 2^h + 2^k)$$

squares the middle entry and halves the right entry g times to get

$$(1,\ a^{2^g},\ 1 + 2^{h-g} + 2^{k-g})$$

multiplies the left entry by the middle entry and decrements the right entry

$$(a^{2^g},\ a^{2^g},\ 2^{h-g} + 2^{k-g})$$

and continues on similarly

$$\begin{aligned}
&(a^{2^g},\ a^{2^h},\ 1 + 2^{k-h})\\
&(a^{2^g+2^h},\ a^{2^h},\ 2^{k-h})\\
&(a^{2^g+2^h},\ a^{2^k},\ 1)\\
&(a^{2^g+2^h+2^k},\ a^{2^k},\ 0),
\end{aligned}$$

and then it returns the first entry, which is indeed a^e.

Fast modular exponentiation is not only for computers. For example, to compute $2^{37} \bmod 149$, proceed as follows,

$$\begin{aligned}
&(1, 2; 37) \to (2, 2; 36) \to (2, 4; 18) \to (2, 16; 9) \to (32, 16; 8)\\
&\to (32, -42; 4) \to (32, -24; 2) \to (32, -20; 1) \to (\boxed{105}, -20; 0).
\end{aligned}$$

2. Fermat Pseudoprimes

Fermat's Little Theorem states that for any positive integer n,

$$\textit{if } n \textit{ is prime then } b^{n-1} = 1 \bmod n \textit{ for } b = 1, \dots, n-1.$$

In the other direction, all we can say is that

$$\textit{if } b^{n-1} = 1 \bmod n \textit{ for all } b = 1, \dots, n-1 \textit{ then } n \textit{ might be prime.}$$

If $b^{n-1} = 1 \bmod n$ for some particular $b \in \{1, \dots, n-1\}$ then n is called a **Fermat pseudoprime base** b.

There are 669 primes up to 5000, but only two values of n (1729 and 2821) that are Fermat pseudoprimes base b for $b = 2, 3, 5$ without being prime. This is a false positive rate of 0.04%. The false positive rate up to 500000 just for $b = 2, 3$ is under 0.01%.

On the other hand, the bad news is that checking more bases b doesn't reduce the false positive rate much further. There are infinitely many **Carmichael numbers**, numbers n that are Fermat pseudoprimes base b for all $b \in \{1, \dots, n-1\}$ coprime to n but are not prime.

Carmichael numbers notwithstanding, Fermat pseudoprimes are reasonable candidates to be prime.

3. Strong Pseudoprimes

The **Miller–Rabin test** on a positive odd integer n and a positive test base b in $\{1, \dots, n-1\}$ proceeds as follows.

- Factor $n-1$ as $2^s m$ where m is odd.
- Replace b by $b^m \bmod n$.
- If $b = 1$ then return the result that n could be prime, and terminate.
- Do the following s times: If $b = n-1$ then return the result that n could be prime, and terminate; otherwise replace b by $b^2 \bmod n$.
- If the algorithm has not yet terminated then return the result that n is composite, and terminate.

(Slight speedups here: (1) If the same n is to be tested with various bases b then there is no need to factor $n-1 = 2^s m$ each time; (2) there is no need to compute $b^2 \bmod n$ on the sth time through the step in the fourth bullet.)

To understand the Miller–Rabin test, consider a positive odd integer n and factor $n-1 = 2^s \cdot m$ where m is odd. Then

$$\begin{aligned}
X^{2^s m} - 1 &= (X^{2^{s-1}m} + 1)(X^{2^{s-1}m} - 1) \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-2}m} - 1) \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-3}m} + 1)(X^{2^{s-3}m} - 1) \\
&\ \ \vdots \\
&= (X^{2^{s-1}m} + 1)(X^{2^{s-2}m} + 1)(X^{2^{s-3}m} + 1) \cdots (X^m + 1)(X^m - 1).
\end{aligned}$$

That is, rewriting the left side and reversing the order of the factors of the right side,

$$X^{n-1} - 1 = (X^m - 1) \cdot \prod_{r=0}^{s-1} (X^{2^r m} + 1).$$

Substitute in any base b,

$$b^{n-1} - 1 = (b^m - 1) \cdot \prod_{r=0}^{s-1} (b^{2^r m} + 1) \bmod n, \quad b = 1, \dots, n-1.$$

If n is prime then $b^{n-1} - 1 = 0 \bmod n$ for $b = 1, \dots, n-1$, and also $\mathbb{Z}/n\mathbb{Z}$ is a field, so that necessarily one of the factors on the right side vanishes modulo n as well. That is, if n is prime then given any base $b \in \{1, \dots, n-1\}$, at least one of the factors

$$b^m - 1, \quad \{b^{2^r m} + 1 : 0 \le r \le s-1\}$$

vanishes modulo n. So contrapositively, if for some base $b \in \{1, \dots, n-1\}$ none of the factors vanishes modulo n then n is composite. Hence the Miller–Rabin test.

A positive integer n that passes the Miller–Rabin test for some b is a **strong pseudoprime base b**. For any n, at least $3/4$ of the b-values in $\{1, \dots, n-1\}$ have the property that if n is a strong pseudoprime base b then n is really prime. But according to the theory, up to $1/4$ of the b-values have the property that n could be a strong pseudoprime base b but not be prime. In practice, the percentage of such b's is much lower. For n up to 500,000, if n is a strong pseudoprime base 2 and base 3 then n is prime.

(Beginning of analysis of false positives.)

Let n be composite. Suppose that n is a strong pseudoprime base b for some b. Then one of the following congruences holds:

$$b^m = 1 \bmod n, \qquad b^{2^r m} = -1 \bmod n \quad \text{for } r = 0, \cdots, s-1.$$

Because $2^s m = n - 1$, any of these congruences immediately implies

$$b^{n-1} = 1 \bmod n,$$

which is to say that n is a Fermat pseudoprime base b.

Next we show that if n is divisible by p^2 for some prime p then there are few bases b for which n is a Fermat pseudoprime base b. In consequence of the previous paragraph, there are thus as few or fewer bases b for which n is a strong pseudoprime base b.

Lemma. *Let n be a positive integer. Let x and y be integers such that n is a Fermat pseudoprime base x and base y,*

$$x^{n-1} = y^{n-1} = 1 \ \mathit{mod}\ n.$$

Let p be an odd prime such that $p^2 \mid n$. If

$$x = y \ \mathit{mod}\ p$$

then

$$x = y \ \mathit{mod}\ p^2.$$

For the proof, first we show that $x^p = y^p \bmod p^2$. This follows quickly from the factorization

$$x^p - y^p = (x - y)(x^{p-1} + x^{p-2}y + \cdots + xy^{p-2} + y^{p-1}),$$

because the condition $x = y \bmod p$ makes each factor on the right side a multiple of p. Second, raise both sides of the relation $x^p = y^p \bmod p^2$ to the power n/p to get $x^n = y^n \bmod p^2$. But because $x^n = x \bmod n$, certainly $x^n = x \bmod p^2$, and similarly for y. The result follows.

Proposition. *Let p be an odd prime. Let n be a positive integer divisible by p^2. Let B denote the set of bases b between 1 and $n-1$ such that n is a Fermat pseudoprime base b, i.e.,*

$$B = \{b : 1 \le b \le n-1 \text{ and } b^{n-1} = 1 \bmod n\}.$$

Then

$$|B| \le \frac{p-1}{p^2} n \le \frac{1}{4}(n-1).$$

To see this, note that the second inequality is elementary to check (to wit, $4(p-1)n \le (p+1)(p-1)n = (p^2-1)n \le p^2 n - p^2 = p^2(n-1)$), so that we need only establish the first inequality. Decompose B according to the values of its elements modulo p,

$$B = \bigsqcup_{d=1}^{p-1} B_d$$

where

$$B_d = \{b \in B : b = d \bmod p\}, \quad 1 \le d \le p-1.$$

For any d such that $1 \le d \le p-1$, if $b_1, b_2 \in B_d$ then the lemma says that $b_1 = b_2 \bmod p^2$. It follows that $|B_d| \le n/p^2$, giving the result.

4. Generating Candidate Large Primes

Given n, a simple approach to finding a candidate prime above $2n$ is as follows. Take the first of $N = 2n+1$, $N = 2n+3$, $N = 2n+5$, ... to pass the following test.

(1) Try trial division for a few small primes. If N passes, continue.
(2) Check whether N is a Fermat pseudoprime base 2. If N passes, continue.
(3) Check whether N is a strong pseudoprime base b as b runs through the first 20 primes.

Any N that passes the test is extremely likely to be prime. And such an N should appear quickly because the slope of the asymptotic prime-counting function is

$$\frac{d}{dx}\left(\frac{x}{\log x}\right) = \frac{\log x - 1}{(\log x)^2} \approx \frac{1}{\log x},$$

so that heuristically a run of $\log x$ gives a rise of 1, i.e., the next prime. And indeed, using only the first *three* primes in step (3) of the previous test finds the following correct candidate primes:

The first candidate prime after	10^{50}	is	$10^{50} + 151$.
The first candidate prime after	10^{100}	is	$10^{100} + 267$.
The first candidate prime after	10^{200}	is	$10^{200} + 357$.
The first candidate prime after	10^{300}	is	$10^{300} + 331$.
The first candidate prime after	10^{1000}	is	$10^{1000} + 453$.

5. Certifiable Large Primes

The **Lucas–Pocklington–Lehmer Criterion** is as follows. *Suppose that p a known prime and that some $N = 1$ mod p is less than p^2 as follows:*

$$N = p \cdot U + 1 \quad \textit{where } p \textit{ is prime and } p > U.$$

Suppose also that there is a base b that suggests that N is prime, in that

$$b^{N-1} = 1 \, mod \, N \quad but \quad \gcd(b^U - 1, N) = 1.$$

Then N is prime.

The proof will be given in the next section. It is just a matter of Fermat's Little Theorem and some other basic number theory. For now, the idea is that the condtions $N = pU + 1$ and $p > U$ say roughly that $p > \sqrt{N}$, while if N is not prime then it has at least one prime factor $q < \sqrt{N} < p$; on the other hand, because $N = 1$ mod p we might hope that all of its prime factors satisfy $q = 1$ mod p and so also $q > p$, contradiction. In general, integers $N = 1$ mod p need not have all their prime factors satisfy $q = 1$ mod p, but the nature of the base b in the LPL criterion ensures that our N does, as we will show.

As an example of using the criterion, start with

$$p = 1000003.$$

This is small enough that its primality is easily verified by trial division. A candidate prime above $1000 \cdot p$ of the form $p \cdot U + 1$ is

$$N = 1032 \cdot p + 1 = 1032003097.$$

And $2^{N-1} = 1 \mod N$ and $\gcd(2^{1032} - 1, N) = 1$, so the LPL Criterion is satisfied, and N is prime. Rename it p.

A candidate prime above $10^9 \cdot p$ of the form $p \cdot U + 1$ is

$$N = p \cdot (10^9 + 146) + 1 = 1032003247672452163.$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{17} \cdot p$ of the form $p \cdot U + 1$ is

$$N = p \cdot (10^{17} + 24) + 1 = 103200324767245241068077944138851913.$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{34} \cdot p$ of the form $p \cdot U + 1$ is

$$\begin{aligned} N = p \cdot (10^{34} + 224) + 1 = & 103200324767245241068077944138854 22 \\ & 468727478629339992494594871028285 13. \end{aligned}$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{60} \cdot p$ of the form $p \cdot U + 1$ is

$$\begin{aligned} N = p \cdot (10^{60} + 1362) + 1 = & 10320032476724524106807794413885422 \\ & 46872747862933999249460892691251842 8 \\ & 80183347221599171194540240682589316 1 \\ & 069777638214340524 34707. \end{aligned}$$

Again $b = 2$ works in the LPL Criterion, so N is prime. Again rename it p.

A candidate prime above $10^{120} \cdot p$ of the form $p \cdot U + 1$ is

$$\begin{aligned} N = p \cdot (10^{120} + 796) + 1 = &1032003247672452410680779441388542\\ &4687274786293399924946089269125184 28\\ \end{aligned}$$

Again $b = 2$ works in the LPL Criterion, so N is prime.

6. Proof of the Lucas–Pocklington–Lehmer Criterion

Our data are

- An integer $N > 1$, presumably large.
- The prime factors q of N, possibly unknown.
- A prime p, to be used to analyze N.

Obviously, if $q = 1 \bmod p$ for each q then also $N = 1 \bmod p$.

The converse does not hold in general. For example, take $N = 10 = 2 \cdot 5$ and $p = 3$. Then $N = 1 \bmod p$ but neither prime factor q of N satisfies $q = 1 \bmod p$.

However, the **Fermat–Euler Criterion** is a partial converse: *Let p be prime. Let N be an integer such that*

$$N = 1 \bmod p.$$

If there is a base b such that

$$b^{N-1} = 1 \bmod N \quad \textit{and} \quad \gcd(b^{(N-1)/p} - 1, N) = 1$$

then

$$q = 1 \bmod p \quad \textit{for each prime divisor } q \textit{ of } N.$$

To prove the Fermat–Euler criterion, let q be any prime divisor of N. Consider the smallest positive integer t such that $b^t = 1 \bmod q$; that is, t is the *order* of the base b modulo q. The set of exponents e such that $b^e = 1 \bmod q$ forms an ideal, making its smallest positive element a generator, which is to say that the exponents e such that $b^e = 1 \bmod q$ are precisely the multiples of t. We will show that $p \mid q - 1$ (i.e., that $q = 1 \bmod p$, the desired conclusion) by showing that t is multiplicatively intermediate to p and $q - 1$.

The Fermat–Euler hypotheses give $b^{N-1} = 1 \bmod q$ and $b^{(N-1)/p} \neq 1 \bmod q$, from which $t \mid N - 1$ and $t \nmid (N-1)/p$, and it follows from these that

$$p \mid t.$$

Also, $b^{q-1} = 1 \bmod q$ by Fermat's Little Theorem, and so

$$t \mid q - 1.$$

Concatenate the previous two displays to get

$$p \mid q - 1.$$

This is the desired result $q = 1 \bmod p$.

The Lucas–Pocklington–Lehmer Criterion builds on the Fermat–Euler Criterion by specializing to the case

$$N = pU + 1, \quad U < p.$$

If such an N satisfies the Fermat–Euler criterion then it must be prime. As already explained, otherwise it has a proper prime factor $q \leq \sqrt{N}$, for which $p \mid q - 1$ by the Fermat–Euler criterion, but the display says that $p > \sqrt{N-1}$ and so $p > \sqrt{N} - 1 \geq q - 1$. The inequality $p > q - 1$ contradicts the condition $p \mid q - 1$, and so no proper prime factor q of N can exist.

Recall the Lucas–Pocklington–Lehmer Criterion:

> *Suppose that $N = pU + 1$ where p is prime and $p > U$. Suppose that there is a base b such that $b^{N-1} = 1 \bmod N$ but $\gcd(b^U - 1, N) = 1$. Then N is prime.*

To prove the criterion we need only verify that the N and p here satisfy the Fermat–Euler criterion, and noting that $U = (N - 1)/p$ does the trick.