Integer factorization in Endymion

José Grimm, Apics Team

June 17, 2005

1 General primality test

The purpose of this section is to give an algorithm that shows that a given integer n is prime, or quasi-prime (i.e. having a high probability to be prime). The function is called `isprime`, it is divided into some sub-functions.

The function **small-int-is-prime** is considered first. Let L be the list 2, 3, 5, 7, ..., 89, 97, the list of all prime numbers between 2 and 100, let P_1 be the product of all prime numbers between 1 and 100, P_2 the product of all prime numbers between 100 and 1000. The first prime after 100 is 101, the first prime after 1000 is 1009, the square of these numbers appear in the algorithm below, which is trivial.

Algorithm 1 (small-int-is-prime) *Argument, a positive integer n. This gives a partial answer to the question: is n prime.*

1. *If $n \in L$, return 'prime'.*
2. *If $\gcd(n, P_1) \neq 1$, return 'composite'.*
3. *If $n < 10201$, return 'prime'.*
4. *If $\gcd(n, P_2) \neq 1$, return 'composite'.*
5. *If $n < 1018081$, return 'prime'.*
6. *Otherwise, return 'maybe'.*

The number P_1 has 37 decimal digits, P_2 has 379 digits. The gcd of two numbers a and b can be computed as follows: let a_0 be the largest, and a_1 be the other one. Let a_{k+1} be the remainder in the division of a_{k-1} by a_k. If the division is exact, the a_k is the gcd, otherwise, continue. A variant of this algorithm produces two numbers u and v such that

$$au + vb = p, \qquad p = \gcd(a, b). \tag{Bezout}$$

Generically, the quotient is small, the cost is linear with the size of n, and $\log n$ divisions are needed. The worst case is when all quotients are 1, case where we consider two consecutive Fibonacci numbers. Thus the cost is $\log^2 n$.

Experiments show that the gcd of n and P_1 costs 30 microseconds when n has 20 digit, 40 microseconds when n has 100 digits. Computing the gcd of n and P_2 costs 41 microseconds if n has 20 digits, and 642 microseconds if n has 380 digits. These timings are very small.

Algorithm 2 (isprime) *Argument, a positive integer n. This gives an answer to the question: is n prime.*

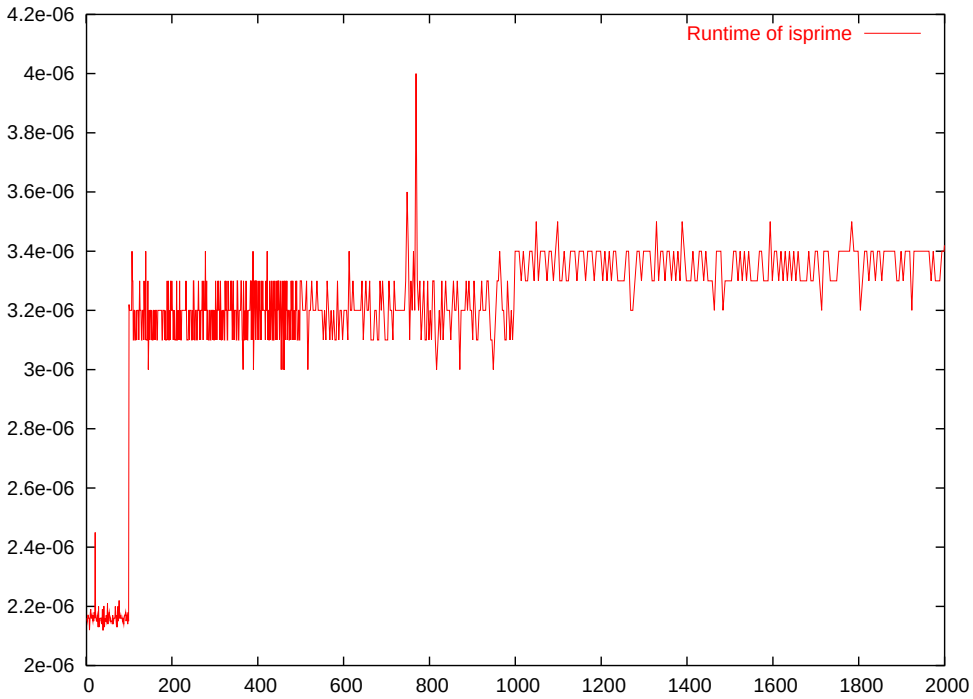

Figure 1: Runtime of isprime, for some small numbers. The maximum is at 769. Limitations of clock precision imply that all measurements are integer multiple of 2.10^{-7}

1. *Call **small-int-is-prime**. This may give an answer.*
2. *Call **Fermat**(n, p) for some small prime numbers p. This may show that n is composite.*
3. *If a definite answer is required, call **true-isprime**.*
4. *Otherwise, let $a_k = 1 + \sqrt{2n/k}$ and $b_k = \sqrt{n/k}$. In each case, the integer part of the quotient is taken, then the integer part of the square root.*
5. *In the case a_k divides n, or b_k divides n, declare n composite. Otherwise, declare n quasi-prime.*

In step 2, the number of tests is controlled by the variable-function **set-max-iter-in-isprime**, the default value being 5. For step 5, the index k in a_k varies between 3 and 9, the index for b_k varies between 5 and 20. These numbers may change in a future version. In the case $n = 21569059132741$, the Fermat test fails five times, and a_4 divides n.

1.1 Cyclicity

Let $Z(n)$ be the set of all integers modulo n, and $G(n)$ the multiplicative group modulo n

$$m \in G(n) \iff 0 < m < n \text{ and } \exists q \; mq = 1 \bmod n \tag{1}$$

We define $\phi(n)$ to be the number of elements of $G(n)$.

Lemma 1 *We have $\phi(n) = n - 1$ if and only if n is prime. This is the negation of: there exists m and q between 1 and $n - 1$ such that $mq = 0 \bmod n$.*

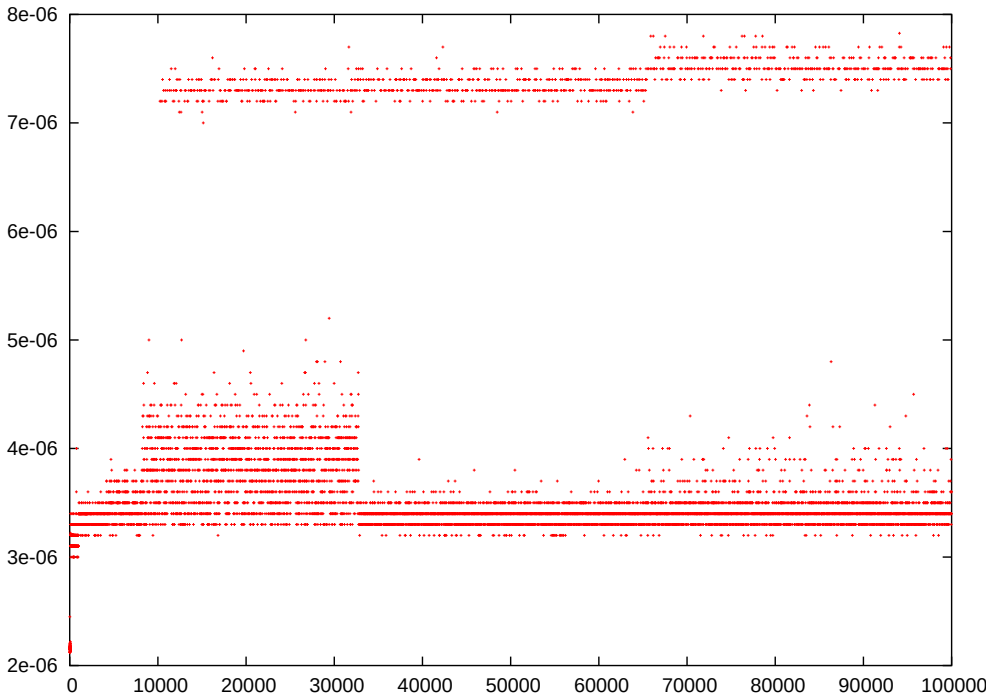

Figure 2: Runtime of isprime, for larger numbers. The numbers we have chosen are: all up to 500, then $502 + 3k$, then $1004 + 5k$, then $2004 + 11k$ then $100069 + 101k$. After that, the interval has been enlarged to 211, 401, 1009. If $n > 10000$, there are integers for which step 4 is required, and you can see the additional cost: it jumps from 3.5 to 7.5 microseconds. A curious phenomenon can be seen: there are lots of numbers less then 2^{15} with a runtime between 3.5 and 4.5 microseconds, and fewer with $n > 2^{15}$. For these, the runtime is in general 3.5 microseconds. As in the previous figure, all timings are integer multiple of 2.10^{-7}.

The lemma is easy: if n is composite, then $n = mq$ for some m and q. If n is prime, and $0 < m < n$, then the Bezout relation gives q such that $mq = 1 + kn$. In the case $mq = 0 \bmod n$, if $0 < m < n$ and $0 < q < n$, there exist numbers $p_1 > 1$ and $p_2 > 1$, p_1 divides m and n, p_2 divides q and n. In particular this implies $m \notin G(n)$. □

Consider the set $\{m, 2m, 3m, \ldots, mn\} \pmod n$. If this set contains n elements, then it contains all numbers between 0 and $n-1$, including 1, and $m \in G(n)$. Otherwise, at least one element is repeated, there is i and j with $m(i-j) = 0$, and $m \notin G(n)$. If $0 < m < n$ we can conclude that n is composite. This gives an awfully bad algorithm: compute the number of elements of this set for every m. The complexity is of the order of n^2. There is a better algorithm: for each m check whether or not m divides n; this algorithm has complexity n. No good algorithm exists. The algorithm shown below assumes that the factorization of $n-1$ can be obtained, this is not always easy.

Assume that

$$n = \prod p_k^{\alpha_k} \tag{2}$$

is the factorization of n into distinct primes. Let $\psi_k(m)$ be $m \bmod p_k^{\alpha_k}$. This gives a function from $Z(n)$ into $Z(p_k^{\alpha_k})$ that preserves addition and multiplication. Let $\psi(m) = (\psi_1(m), \psi_2(m), \ldots)$.

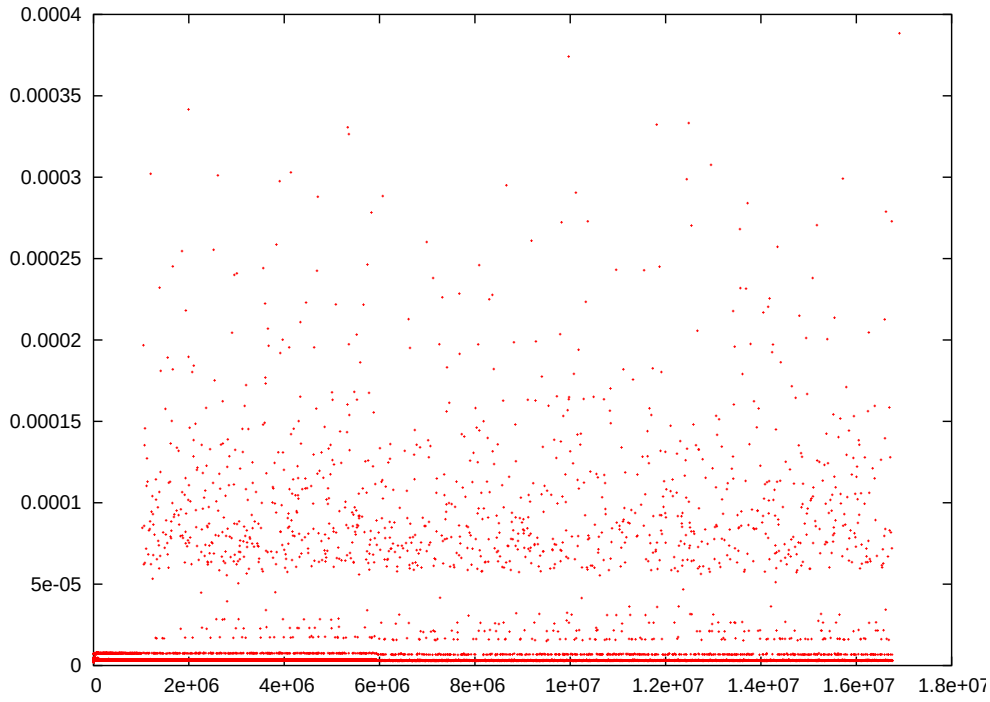

Figure 3: Runtime of isprime, for larger numbers.

We get an application

$$\psi : Z(n) \longrightarrow \prod Z(p_k^{\alpha_k}). \tag{3}$$

The Chinese Remainder Theorem says that this is a bijection, preserving addition and multiplication (the theorem is just a consequence of the Bezout relation). In particular, it maps invertible elements to invertible elements, hence induces a function

$$\psi : G(n) \longrightarrow \prod G(p_k^{\alpha_k}). \tag{4}$$

As a consequence, $\phi(n) = \prod \phi(p_k^{\alpha_k})$. This can be restated: if a and b are coprime, then $\phi(a)\phi(b) = \phi(ab)$. Note that $m \in G(p^\alpha)$ if and only if m is not a multiple of p, there are $p^{\alpha-1}$ multiples of p between 1 and p^α, so that $\phi(p^\alpha) = p^\alpha(1 - 1/p)$ hence

$$\phi(n) = \prod (p_k^{\alpha_k} - p_k^{\alpha_k - 1}) = n \prod (1 - 1/p_k). \tag{5}$$

Lemma 2 *The following relation holds for any integer* n

$$\sum_{d|n} \phi(d) = n. \tag{6}$$

Proof. Assume first $n = p^k$, and let $d_i = p^i$. The only divisors of n are the d_i. We have $\phi(d_0) = d_0$, and $\phi(d_i) = d_i - d_{i-1}$ otherwise. Hence the sum is d_k, thus n.

Otherwise, factor n as in (2). Then d divides n if and only if $d = \prod d_i$, where each d_i divides $p_i^{\alpha_i}$. Use $\phi(d) = \prod \phi(d_i)$. Then $\sum \phi(d)$ is the product of the $\sum \phi(d_i)$. By what precedes, this sum is $p_i^{\alpha_i}$, and the product is n. □

We say that a group G is of *order* n if it has n elements; we say that an element x is of order n if the group it generates (formed of all x^i) is of order n. If G is the group generated by x, then G is called *cyclic*, and x is a *generator*. Consider a group G, an element x. Let $C(y)$ be the set of all yx^k. In general, this is not a group, but its number of elements is independent of y, it is the order of x. Suppose that $C(y)$ and $C(z)$ intersect; there exist α and β such that $yx^\alpha = zx^\beta$. From this we deduce $C(y) = C(z)$. Thus G is the disjoint union of all $C(y)$, its order is a multiple of n.

This can be restated as: If $x \in G$, then the order of x divides the order of G. We can also say: if g is the order of G then $x^g = 1$, whatever x. This equation can also be satisfied by a number smaller than g, for instance, in $G(8)$, that has four elements, we have $x^2 = 1$. Of course, G is not cyclic. Note that, if x is of order n, then x^m is of order $n/\gcd(n,m)$, which is obvious from the Bezout relation, so that teh set $\{x^i\}$ contains $\phi(n)$ elements of order n. If p is prime, then every x between 1 and $p-1$ is in $G(p)$, thus has an order that divides p. Thus

Theorem 1 (Fermat) *If p is prime, $0 < x < p$, then $x^{p-1} = 1 \pmod p$.*

Lemma 3 *If n is prime, then $G(n)$ is cyclic.*

Let's consider a prime number p. Let $\psi(d)$ be the number of elements of $G(p)$ whose order are exactly d. We pretend $\psi(d) \leq \phi(d)$. This is obvious if no element is of order d. Otherwise, let's consider one x of order d. If $y = x^k$, then $y^d = 1$. There are d distinct powers of x, hence d solutions to $X^d - 1 = 0$. Since p is prime, $Z(p)$ is a field, and the polynomial $X^d - 1$ has not more than d roots. Thus, if y is a generator, it must be a power of x. Since the powers of x cainain $\phi(d)$ elements of order d, we have: $\psi(d)$ is zero or $\phi(d)$. If $n = p - 1$ we have

$$\sum_{d|n} \psi(d) = n. \tag{7}$$

This equation is equivalent to say that every element of $G(p)$, a group with n elements, has an order that divides n. If we compare with (6), we see that $\phi(d) = \psi(d)$, whenever d divides n. This is in particular true if $d = n$, i.e. if $d = p - 1$; this shows that $G(p)$ has at least one generator, in fact, it has $\phi(p-1)$ generators. □

Examples: $G(2)$ is trivial, it contains only 1; the group $G(3)$ has two elements, 1 and 2, the square of 2 is 1, and 2 is a generator. If n is an odd prime, $n = 1 + q2^k$, q odd, then $\phi(n-1) = 2^{k-1}\phi(q)$. This is even if $k > 1$ or if $q > 1$ (if q is odd, $q > 1$, then $\phi(q)$ is even). Thus, if $n > 3$, $\phi(n-1)$ is even, there are an even number of generators (if fact, if x is a generator, so is its inverse).

We may ask the following questions: are there other cyclic groups. As said above, $G(8)$ not cyclic. This is because, if m is odd, we have $m = 2k + 1$ and $m^2 = 4k(k+1) + 1$, and $k(k+1)$ is even. Hence

$$\forall m \text{ odd }, \exists k \qquad m^{2r} = 1 + 8rk$$

for $r = 1$. If we take squares, we see that this is true whenever r is a power of two. Thus $m^{2r} = 1$ whenever $m \in G(8r)$. Since $G(8r)$ has $4r$ elements, it cannot be cyclic. For the case of $G(p^k)$ with p odd we start with:

Lemma 4 *Let p be some prime number. Let $f(n)$ be the power of p in $n!$, and $g(n,i)$ the power of p in $\binom{n}{i}$. In the case where $n = p^k$ and $i = qp^j$, with q coprime to p, we have $g(n,i) \geq k - j$.*

We have

$$f(n) = \lfloor \frac{n}{p} \rfloor + \lfloor \frac{n}{p^2} \rfloor + \ldots + \lfloor \frac{n}{p^k} \rfloor + \ldots \tag{8}$$

In the case $n = mp^j$, we can write this as

$$f(n) = m(1 + p + \ldots + p^{j-1}) + f(m) = mP_j + f(m).$$

We have $g(n,i) = f(n) - f(i) - f(n-i)$. If n and i are as in the lemma, $m = p^{k-j}$, the factors of P_j are $m - q - (m - q)$. Thus

$$g(n,i) = g(m,j).$$

This means that we need only to prove the theorem in the case $j = 0$. We have

$$b\binom{a}{b} = a\binom{a-1}{b-1}$$

so that, if a and b are coprime, $\binom{a}{b}$ is a multiple of a. □ We continue with:

Lemma 5 *Let p be an odd prime, $k \geq 2$, $q = p^{k-2}(p-1)$. Let θ be an integer such that, if $x = \theta^{p-1} - 1$, then x is zero modulo p, but not modulo p^2. Then θ^q is not 1 modulo p^k.*

We start with

$$\theta^q = (1+x)^{p^{k-2}} = \sum_i \binom{p^{k-2}}{i} x^i.$$

The first term is one, the second is xp^{k-2}, this is not zero mod p^k. Remaining terms are zero. This is because the power of p is at least $g(p^{k-2}, i) + i$. According to the previous lemma, this is at least $k - 2 + i - j$. Thus, we have to show that $i \geq j + 2$. This is true for $j = 0$, since we consider only $i \geq 2$. Otherwise, we know $i = qp^j$, so that it suffices to show $p^j \geq j + 2$. For $j = 1$, this gives $p \geq 3$ (remember that p is odd). For $j \geq 2$, we have $2j \geq j + 2$, and $p^j \geq 2j$, which is true if $p \geq 2$ and $j \geq 2$. □

Lemma 6 *If p is an odd prime, then $G(p^k)$ is a cyclic group.*

In fact, consider ξ, a generator modulo p. Let θ be as follows. If $\xi^{q-1} \neq 1 \pmod{p^2}$, we take $\theta = \xi$, otherwise $\theta = \xi + p$. Then θ is a generator modulo p, and the conditions of the previous lemma are satisfied. What is the order modulo p^k? it is some divisor of the order of the group, namely $g = p^{k-1}(p-1)$. If the order is not g, then there is a prime r such that $\theta^{g/r} = 1 \pmod{p^k}$. Assume first that r divides $p - 1$. Then $\theta^{g/r} = 1 \pmod{p}$. However $\theta^p = \theta \pmod{p}$, hence $\theta^{(p-1)/r} = 1 \pmod{p}$, and this contradicts the fact that θ is a generator modulo p. Hence the order of θ is a multiple of $p - 1$, is it $p^\alpha(p-1)$. We have $\alpha \geq k - 1$, so that the order is g, since otherwise this would imply $\theta^q = 1$. □

Note that if p is an odd prime, then $G(2p^k)$ is a cyclic group. This is really because $G(p^k)$ and $G(2p^k)$ are isomorphic: consider f such that $f(m) = m$ is m is odd, and $f(m) = m + p^k$ if m is even. Then f is a bijection from $G(p^k)$ to $G(2p^k)$. It preserves multiplication.

Lemma 7 *Let G be a commutative group, a_1 and a_2 in G with order g_1 and g_2. Let α be the gcd of g_1 and g_2. Then the order of $a_1 a_2$ divides $g_1 g_2/\alpha^2$ and multiplies $g_1 g_2/\alpha$. In particular, if g_1 and g_2 are coprime, the order is $g_1 g_2$.*

In fact, let $g_1 = \alpha h_1$ and $g_2 = \alpha h_2$. We have $(a_1 a_2)^{\alpha h_1 h_2} = 1$. On the other hand, if $(a_1 a_2)^\gamma = 1$, raising thius to the power αh_1 gives $a_2^{\alpha h_1 \gamma} = 1$, so that $\alpha h_1 \gamma$ is a multiple of αh_2 and γ is a multiple of h_2. Thus γ is a multiple of $h_1 h_2$. □

Lemma 8 *If G is a commutative group, the product of some groups, G_1, G_2, etc., then G is cyclic if and only if each G_i is cyclic, and the orders of the G_i are coprime.*

If suffices, by induction, to prove the lemma in the case of two factors. If $x = (x_1, x_2) \in G$, we have $x = a_1 a_2$, where $a_1 = (x_1, 1)$, and $a_2 = (1, x_2)$. If G_1 is of order g_1, and G_2 of order g_2, then $a_1 a_2$ is of order at most the gcd of g_1 and g_2. If we want G to be cyclic, and x a generator, since G has $g_1 g_2$ elements, we need $g = g_1 g_2$, i.e. g_1 and g_2 coprime. Moreover, if x is a generator, then x_1 is a generator of G_1 and x_2 of G_2. These conditions are sufficient. □

Consider now an integer n, and (2) its factorization. If p_k is odd or $p = 2$ and $\alpha_k > 1$, then $\phi(p_k^{\alpha_k})$ is even. If n has two such factors, $G(n)$ cannot be cyclic. We have shown:

Theorem 2 *The group $G(n)$ is cyclic if and only if $n = p^k$, or $n = 2p^k$ or $n = 1$ or $n = 4$, where p is an odd prime.*

Algorithm 3 (generator-modp) *Argument, an integer n. This returns a generator modulo n, if one exists.*

1. *If $n = 2$, return 1, if $n = 4$, return 3.*
2. *Factor $n = \prod p_i^{\alpha_i}$, with $p_1 < p_2 < \ldots < p_k$.*
3. *If there are more than two factors, return 'failed'.*
4. *If $k = 2$ (case of two factors), if $p_1 \neq 2$, or $\alpha_1 > 1$, return 'failed'. Set p to p_2, and e to α_2.*
5. *If there is one factor, set p to p_1, and e to α_1. Return 'failed' if $p_1 = 2$.*
6. *Compute g, a generator modulo p via **igenerator**(p).*
7. *If $e \geq 2$, and $g^{p-1} = 1 \pmod{p^2}$, replace g by $g + p$.*
8. *If n is even, and g is even, replace g by $g + p^k$.*
9. *Return g.*

1.2 Primality tests

Let's consider the following three algorithms. Algorithmn A is: For k beween 1 and $n - 1$, compute the remainder of k by n. Returns true if false is found. Algorithm B depends on x, it is: set $y = x$, for k between 1 and $n - 1$, replace y by the remainder in the division of xy by n. For each integer i between 1 and $n - 1$, count how many times it is a y. Return true if all these counts are 1. Algorithm C is the same, but we count only the the appearance of 1.

If the algorithm returns true, then n is prime. If algorithm A returns false then n is composite. For the other algorithms, we must test a lot of numbers x. Algorithm B cannot be used in practice, since it needs a huge table. All algorithms have a runtime proportional to n. The only efficient algorithm we know assumes that the factorization of $n - 1$ is known. This is because $x^k \neq 1$ for $0 < k < n - 1$ is equivalent to $x^k \neq 1$ whenever $k = (n - 1)/r$, for any prime r. For instance, if $n = 17$, this is $x^8 \neq 1$. Three multiplications are required. Half of the numbers satisfy this equation.

This condition is not enough: we must add $x^{n-1} = 1$, a relation that holds for any x, not only for generators. For instance, if $n = 6$, we have $2^5 = 2$. In fact, we have $2^3 = 2$, and $2 \notin G(6)$.

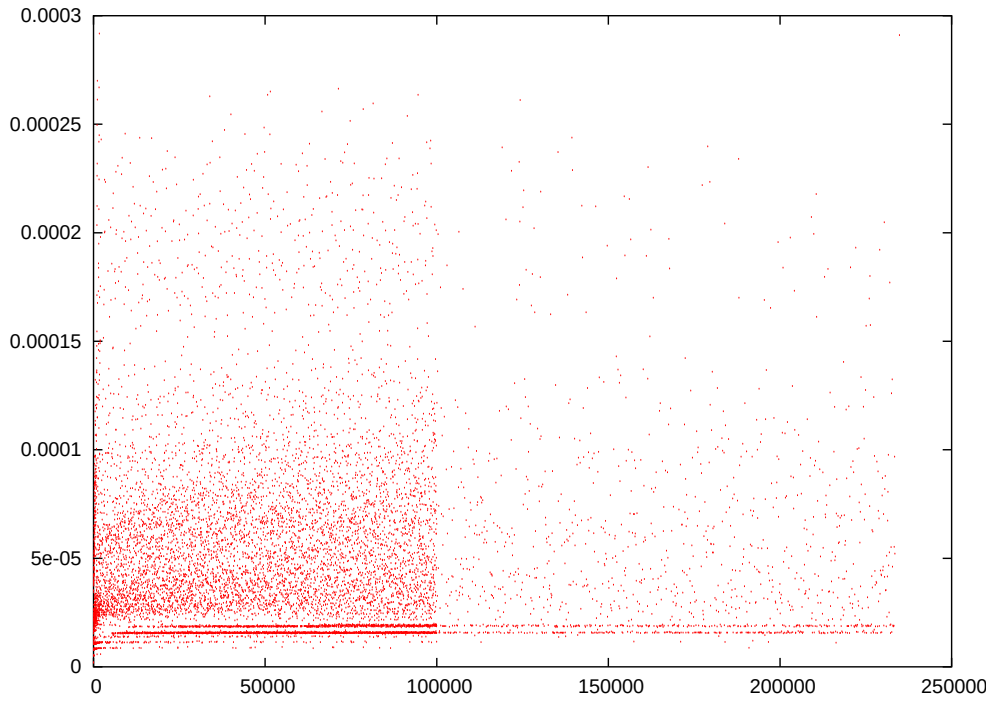

Figure 4: Runtime of 'generator-modp'. For $n < 100000$, one number out of 11 is chosen, otherwise one out of 101.

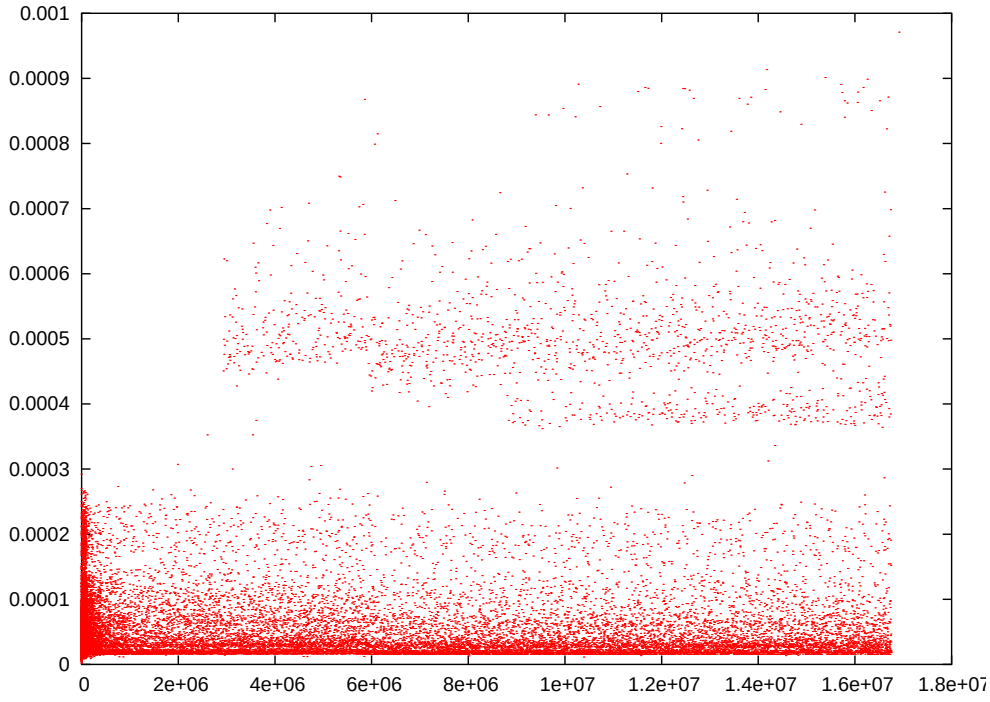

Figure 5: Runtime of 'generator-modp', for larger numbers.

The complexity of the algorithm is the following. Computing x^k costs $\log k$ multiplications. If N is the size of n, each multiplication costs N^2. We can estimate $k \leq n$, hence $\log k \leq N$. The number of prime factors is also bounded by $\log n$, so that the cost is N^4. This is something reasonable; the question is however: how many integers x do we need to try? We shall give a variant, where some numbers are tested; each test costs N^3, and will help to prove that n is prime (or show that n is composite). For each prime r, dividing $n-1$, some numbers have to be tested, but we have no idea of how many.

Let's consider a group with g elements, and the factorisation of g.

$$g = \prod p_k^{\alpha_k}. \tag{9}$$

If the group is cyclic, we have a generator, hence for each i an element a_i such that

$$a_i^g = 1 \qquad a_i^{g/p_i} \neq 1. \tag{10}$$

It suffices to chose $a_i = x$. On the other hand, assume that these relations are true. Let $q_i = g/p_i^{\alpha_i}$, and $b_i = a_i^{q_i}$. Then (10) is equivalent to

$$b_i^{p_i^{\alpha_i}} = 1 \qquad b_i^{p_i^{\alpha_i - 1}} \neq 1. \tag{11}$$

In other words, b_i is of order $p_i^{\alpha_i}$, and $b = \prod b_i$ is of order g. This gives a method for finding a generator.

Algorithm 4 (igenerator) *Argument, a number n, optionally, a list L. If L is given, it is the list of factors of $n-1$, some elements being marked. In this cases, steps 1 and 2 are useless. This returns a generator of $G(n)$, or fails if n is composite.*

1. *Factor $n-1$ as $\prod p_i^{\alpha_i}$.*
2. *Initialise g to 1. Unmark all positions i.*
3. *Repeat the following steps, until all primes are marked, where P is 2, 3, 5, etc, a prime number.*
4. *Consider all positions i that are not marked. We have a prime p, an exponent k.*
5. *Let $B = P^{(n-1)/p} \pmod n$. If $B^p \neq 1 \pmod n$, return 'composite'. If the gcd of $B-1$ and n is neither 1 nor n, return 'composite'. If the gcd is not n, mark the position i.*
6. *Let $q = (n-1)/p^k$, $r = p^{k-1}$, $b = P^q \pmod n$. If $b^r \neq 1 \pmod n$, multiply g by b, and mark i.*
7. *If all positions are marked, then n is prime and a generator of $G(n)$ is g.*

Some comments. If we want to find a generator of $G(n)$, we know that n is prime, and we can skip step 5. If we want to prove that n is prime, we can skip the computation of g in 6. We have $B = b^r$. Conditions (11) are $b^r \neq 1$ and $b^{rp} = 1$. This last condition is true, if we assume n prime. It can be used in 5 to show that n is composite. In the case n prime, either $B = 1$, case where the gcd of $B-1$ and n is 1, or $B \neq 1$, case where the gcd is not n, hence is 1. If n is composite, the gcd could be a proper factor, and this is tested. Said otherwise, we replace condition (11) by the stronger one

$$a^{n-1} = 1 \pmod n \qquad \gcd(a^{(n-1)/p_i} - 1, n) = 1. \tag{12}$$

Lemma 9 *If* (12) *is true for some a, it is also true for some prime not greater than a. This explains why, in step 3 of the algorithm, we chose only prime numbers P.*

In fact, assume $a = bc$. If $b^s = 1$ and $c^s = 1$, then $(bc)^s = 1$. The converse can be false. Hence, we can miss an opportunity to show that n is composite. Let $q = (n-1)/p_i$. If $b^q - 1$ is coprime to n, condition (12) is true for b. If b^q is 1 mod n, then (12) is true for a if and only if it is true for c. Otherwise n is composite, and can can miss an opportunity to show it. □

The reason why we introduce the gcd is the following (Knuth exercise 4.5.4.26).

Lemma 10 *Assume $n = 1 + fr$, $0 < r \leq f + 1$. If* (12) *is true for every prime factor p_i of f, then n is prime.*

Let's write $n = 1 + QR$, where Q is coprime to R, R has the same prime factors as f, so that $f \leq R$. Write $b = a^Q$. Then

$$b^R = 1 \pmod{n} \qquad \gcd(b^{R/p_i} - 1, n) = 1. \tag{13}$$

Let P be a prime factor of n. Then $b^R = 1 \pmod{P}$, and $b^{R/p_i} \neq 1 \pmod{P}$. If for each prime factor of R we have such a b, this gives an element of order R modulo P. Thus, $R < P$. In particular $f < P$. Suppose n composite. Since all pime factors are greate then f we have $n \geq (f+1)^2 \geq r(f+1) = rf + r = n - 1 + r$. This implies $r = 1$. But $r = 1$ says $f = n - 1$, and we have a generator modulo n.

Consider $n = 31$. We have to find numbers such that $a^{15} \neq 1$, $a^{10} \neq 1$, $a^6 \neq 1$. There are 15 numbers satisfying $y^{15} = 1$, (this condition is equivalent for y to be a square modulo 31). These numbers are 1, 2, 4, 5, 7, 8, 9, 10, 14, 16, 18, 19, 20, 25, and 28. There are nearly $\sqrt{n/2}$ integers between $n/2$ and n that are a square (here 16 and 25). It could happen that all numbers less then $n/2 - \sqrt{n/2}$ are square mod n. In this case, the generators are 3, 11, 12, 13, 17, 21, 22, 24, 25 and 28. Note: we have $2^5 = 1$, so that $2^{15} = 1$. Thus 2 is not a generator. However, since $2 - 1$ is coprime to 31, the relation $2^5 = 1 \pmod{31}$ shows that any prime factor of 51, has order at leat 5, hence is at least 6. Since $31 < 6^2$ it is prime.

Algorithm 5 (true-isprime) *Argument, a number n. This checks whether n is prime or not*

1. *Factor $n - 1 = \prod p_i^{\alpha_i}$.*
2. *Write $n - 1 = fr$, with $r = 1$. For each i, write $r' = rp_i^{\alpha_i}$ and $f' = f/p_i^{\alpha_i}$. If $r' \leq f' + 1$, replace r by r' and f by f', and mark the index i.*
3. *Call **igenerator**, with arguments the prime decomposition with the marks.*

Note. In **igenerator**, we have a test: $B^p \neq 1 \pmod{n}$. This is the same as $P^{n-1} \neq 1 \pmod{n}$; we execute it only once for each prime P. Moreover, we skip the test for all primes that satisfy the Fermat criterion.

1.3 Fermat test

We shall use algorithm P of Knuth (section 4.5.4).

Algorithm 6 (Fermat) *Arguments n, p. This procedure may find that n is composite.*

1. *If this is not already done, write $n = 1 + x2^k$, where x is odd.*
2. *Consider $a_0 = p^x \pmod{n}$.*

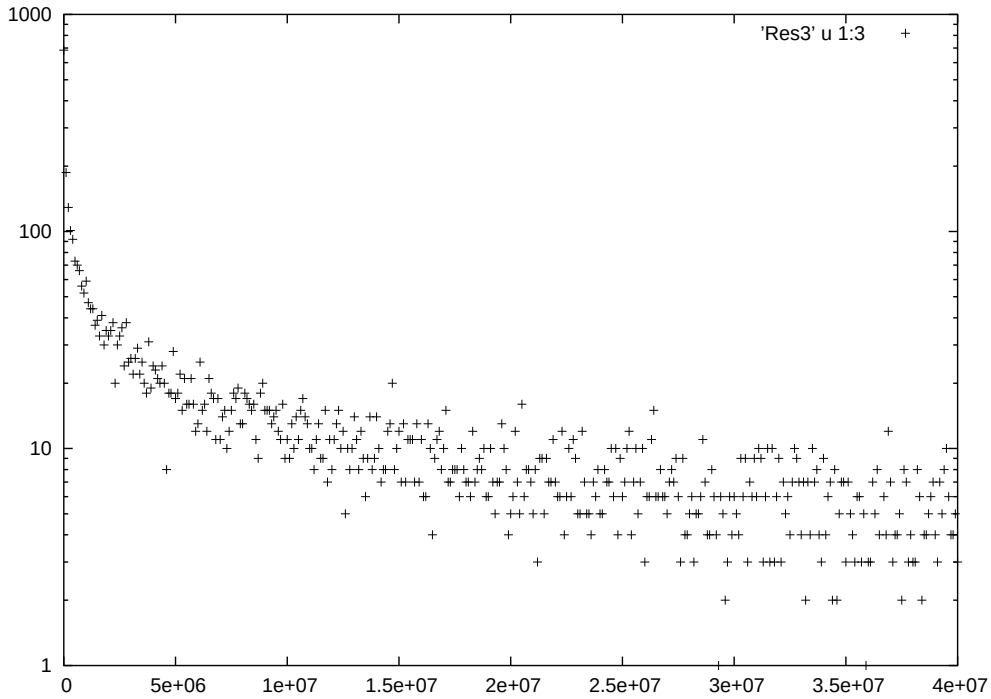

Figure 6: This figure shows, for every n, the number of times $Q(k) \leq 100$ where k is between n and $n + 10^5$, k is odd and composite.

3. *If $a_0 = 1$, return 'maybe'.*
4. *Consider $a_i = a_{i-1}^2 \pmod n$. Return 'composite' if $a_i = 1$ or 'maybe' if $a_i = -1$ mod n. Do this check for $i = 0, 1, 2, \ldots$.*
5. *Do not compute a_k, return 'composite'.*

The algorithm returns 'composite' in two cases. In the case where $a = b^2$ and $a = 1 \pmod n$, we have $(b-1)(b+1) = 0 \pmod n$. Thus either $b = 1$, or $b = -1$, or there is a non-trivial factor between n and $b - 1$. What is a_k? this is $x^{n-1} \mod n$. In the case $a_k = 1$, we can conclude that n is composite (as before), otherwise, we can conclude that n is composite (Fermat Theorem).

Theorem 3 *If algorithm **Fermat** fails to detect that n is prime for a random number p, then the probability that n is composite is less than $1/4$. In fact, it is much less than that, see figures.*

Of course, we do not call the algorithm with a random number, but with a small prime. We shall assume n odd (in fact, the numers we try have no factor less 100).

Some notations. We factor $n = \prod q_i^{e_i}$. The group $G(q_i^{e_i})$ is cyclic, let ξ_i be a generator, let $B_i = (q_i - 1)q_i^{e_i - 1}$ be its order. The Chinese Remainder Theorem says that $a \in G(n)$ is uniquely characterised ny all remainders mod $q_i^{e_i}$, hence the numbers $r_i(a)$ such that

$$a = \xi_i^{r_i(a)} \pmod{q_i^{e_i}}.$$

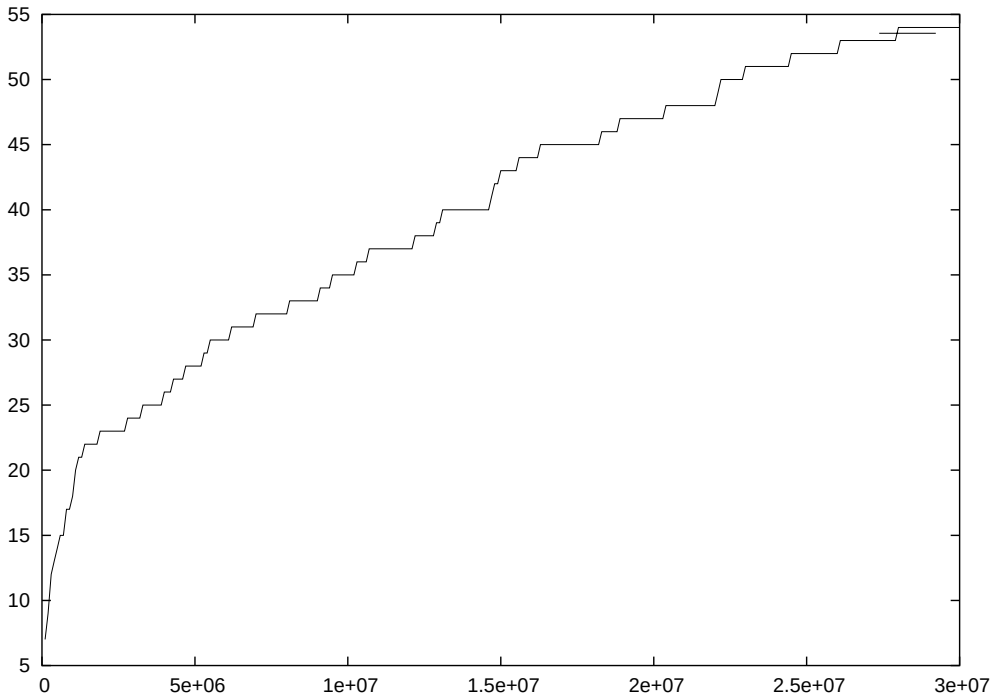

Figure 7: This figure shows, for every n, the number of times $Q(k) \leq 5$ where k is odd and composite and $\leq n$.

Write $n = 1 + x2^k$, $q_i = 1 + x_i 2^{k_i}$, where x and x_i are odd. Write $x'_i = \gcd(x, x_i)$. We have

$$\gcd(x2^j, B_i) = \gcd(x2^j, (q_i - 1)q_i^{e_i - 1}) = \gcd(x2^j, q_i - 1) = \gcd(x2^j, x_i 2^k) = 2^j \gcd(x, x_i) = 2^j x'_i,$$

provided $0 \leq j \leq k$, because $q_i \neq 2$ since n is odd, x and x_i are odd, q_i does not divide x. The factors q_i are sorted such that $k_1 \leq k_2 \leq \cdots \leq k_s$. Let $K = k_1$. We have $q_i = 1 \pmod{2^K}$, hence $n = 1 \pmod{2^K}$, in other words, $K \leq k$.

Define b_n as the number of integers for which the algorithm fails (i.e. returns 'maybe'). These are numbers a such that either $a^x = 1$, or $a^{x2^j} = -1$ for some $j > 0$.

The first case of failure is $a^x = 1$. This is equivalent to $a^x = 1$ mod each $q_i^{e_i}$. This is equivalent to $r_i(a)x = 0 \pmod{B_i}$. The number of integers r that satisfy $rx = 0 \pmod{B}$ is $\gcd(x, B)$, here $\gcd(B_i, x) = x'_i$. Hence, the number of solutions to $a^x = 1$ is hence $\prod x'_i$.

Consider now the number of solutions to $a^{x2^j} = -1$. The condition is now

$$r_i(a)x2^j = B_i/2 \pmod{B_i}.$$

The equation can be written as $rx2^{j+1} = (2l + 1)B_i$ for some l. Remember that the power of 2 in B_i is k_i, so that the equation has no solution of $j + 1 > k_i$. In particular if $j \geq K = k_1$, there is at least one i for which there is no solution. Otherwise, $j < k_i$, there is a solution r', all other solutions satisfy $(r - r')x2^j = 0 \pmod{B_i}$, and the number of solutions is $\gcd(x2^j, B_i) = x'_i 2^j$. Since we have s factors, the total number is $2^{js} \prod x'_i$. Hence

$$b_n = (1 + \sum_{j=0}^{K-1} 2^{js}) \prod x'_i. \tag{14}$$

Let

$$P(n) = b_n/n \qquad Q(n) = 1/P(n) \tag{15}$$

The quantity P is the probability of failure, the theorem claims $Q \geq 4$. Note that Q is independent of e_i, so that if $m = \prod q_i$, we have

$$Q(n) = Q(m) \prod p_i^{e_i - 1}$$

As a consequence, when n has repeated factors, the probability of failure becomes smaller. In particular, if we assume that n has no prime factor less than 1000, this gives $Q \geq 1000$.

Let A be the first factor in b_n. This is also

$$A = 1 + \frac{2^{Ks} - 1}{2^s - 1}. \tag{16}$$

Introduce $B = 2^{Ks}$. We claim that $B/A \geq 2^{s-1}$. If $y = 2^{-s}$, this is equivalent to

$$w_K \leq 2 \text{ where } w_K = \frac{1 - y^K}{1 - y} + y^{K-1}.$$

We have $w_k = w_{k-1} + (2y - 1)y^{k-2}$, $w_1 = 2$, and $2y - 1 \leq 0$, thus the claim.

Since K is the smallest ok the k_i, the quantity $r/B = \prod 2^{k_i - K}$ is an integer. Let's introduce $\phi(n)$. We have $\phi(n) = r \prod x_i \prod \prod q_i^{e_i - 1}$, hence

$$Q = \frac{n}{\phi(n)} \frac{\prod x_i}{\prod x_i'} \frac{r}{B} \frac{B}{A} \prod q_i^{e_i - 1}.$$

In this product all factors are ≥ 1. We want to show $Q \geq 4$, and that most of the time $Q \geq 8$. Because $B/A \geq 2^{s-1}$, we have $Q \geq 8$ if $s > 3$, so that we need only consider the case of one, two and three factors.

Consider the case when n has a single prime factor. We have $b_n = \gcd(2^K x, q_1 - 1)$. In particular $b_n \leq q_1$, and $Q \geq q_1^{e_1 - 1}$. In the case $n = 9$, we have $Q = 9/2$, because $b_9 = 2$ (The Fermat criterion fails for $a \pm 1$, so that $A \geq 2$, in this case we have equality). In all other cases where n is composite, odd and the power of a prime $n = p^e$, we have have either $p \geq 5$ or $e \geq 3$, hence $Q \geq 5$ or $Q \geq 9$. In fact, assume that n has no prime factor less than 1000. Then $Q \geq 1000$.

We consider now the case $s = 2$ or $s = 3$. Then B/A is at least 2 or 4. In fact can be divided by a power of a prime p, then $Q \geq 2p$ or $Q \geq 4p$, since $p \geq 3$, this will show the theoreom. Hence the only non-trivial case is when n is the product of two distinct primes, say

$$n = 1 + x2^k = (1 + y2^{k_1})(1 + z2^{k_2}).$$

If $k_2 = k_1 + \delta$, then

$$\frac{r}{A} = \frac{3.2^\delta.2^{2k}}{2 + 2^{2k}} = 2^\delta \frac{3}{1 + 2^{1-2K}} \geq 2^{\delta+1}.$$

We have

$$Q = \frac{1 + y2^{k_1}}{y'2^{k_1}} \frac{1 + z2^{k_2}}{z'2^{k_2}} \frac{r}{A}. \tag{17}$$

If $\delta \geq 2$, r/A is at least 8. In any case, it is at least 2. Consider $\delta = 0$ first. We have

$$x2^{k-K} = y + z + yz2^K$$

and y' is the gcd of x and y. If $w = y/y'$, the first factor in (17) is at least w, so that, if w is not trivial, we have $Q \geq 6$. (remember that y is odd). If $w = 1$, this says that y divides x. The previous equation implies that y divides z. Same conclusion if we exchange the roles of y and z. Note that we have $y \neq z$, because n has two distinct prime factors. Thus, if neither y divides z, nor z divides y, we have $Q \geq 18$. In the case

$$n = (1 + y2^k)(1 + yz2^k) \qquad \text{(two prime factors)}$$

we have $Q \sim 2z$.

Consider finally the bad case, where $\delta = 1$. If $K \geq 2$, we have $r/A \geq 16/3$, and if $K = 1$ we have $r/A = 4$. Nothing more can be gained if y and z divide x; this condition implies $x = y = z$, and $n = (1+2y)(1+4y)$. According to Knuth, the least such number is obtained for $y = 24969$. □

In the case $s = 3$, we have $B/A \geq 4$. This proves the theorem. We pretend however that a better bound can be found. Obviously, we exclude the case $x_i' \neq x_i$ (a factor 3 can be gained here). A factor 2 can be gained in r/B if $k_i \neq K$. The situation is now

$$n = (1 + x_1 2^K)(1 + x_2 2^K)(1 + x_3 2^K) = 1 + x2^k.$$

Expanding, and looking at powers of 2 shows that $K = k$. The condition $x_1' = x_1$ implies that x_1 divides x. Since $x = x_2 + x_3 + x_2 x_3 2^K \pmod{x_1}$, we have $(1+x_2 2^K)(1+x_3 2^K) = 1+x_1 y_1 2^K$ for some y_1, hence $n = (1 + x_1 2^K)(1 + x_1 y_1 2^K)$. Moreover, two other relations, with indices 2 and 3 also hold. I do not know if such numbers exist.

1.4 Debugging

This chapter is a test. Its aim is to find numbers for which the Fermat criterion fails. Using the procedures defined below, we found the numbers shown here. The value n is the number of primes needed to make the Fermat test work.

$n = xy$	x	y	p	n
118670087467	$1 + 2p$	$1 + 8p$	86121	5
315962312077	$1 + 4p$	$1 + 16p$	70263	5
354864744877	$1 + 4p$	$1 + 16p$	74463	5
602248359169	$1 + 2p$	$1 + 10p$	173529	5
457453568161	$1 + 4p$	$1 + 12p$	97623	5
307768373641	$1 + 8p$	$1 + 16p$	49035	5
528929554561	$1 + 4p$	$1 + 12p$	104973	5
546348519181	$1 + 4p$	$1 + 8p$	130665	5
11377272352951	1686511	4x-3	p	N
22749134240827	2384803	4x-3	p	N

1.5 Algorithm $p-1$ of Pollard

I do not know exactly how this works. The idea is the following. Let $s = s_0$ be the seed, a number coprime with n. Let R be some integer, and compute $s^R - 1$. If R is the product of the r_i's, we do this by computing $s_i = s_{i-1}^{r_i}$. Let k denote the number of factors of R. If $\gcd(s^R - 1, n)$ is one, we are not lucky. Since for each prime p dividing n we have $s^{p-1} = 1 \mod p$, this means that for each p dividing n, $p-1$ does not divide R. So, if we are not lucky, then n has no small primes.

Assume now that we are lucky, in other words, there exists a least i such that $\gcd(s_i - 1, n)$ is not 1. If moreover this is not n, then we have a nontrivial factor of n. If this is n, we may try to factor r_i. Assume first that r_i is $p_1 \dots p_m$, where each p_j appears with exponent 1. We define $S_0 = s_{i-1}$ and $S_j = S_{j-1}^{p_j}$, so that $S_m = 1 \mod n$. This equation means that there is a least j such that $\gcd(S_j - 1, n)$ is not one. If this gcd is not n, then we have a nontrivial factor of n. Otherwise, we start again with s^{p_j} as seed instead of s. In this case, we will get $S_{j-1} = 1$. This equation means that the second time, we shall execute less steps than the first time. The algorithm will finish if we decide to stop in the case where $i = 1$. We return 1 in this case. We return 0 in case $\gcd(s^R - 1, n) = 1$. In all other cases, we return a nontrivial factor.

1.6 Algorithm of Morrisson and Brillhart

Assume that n is not a perfect square, and consider g, the integer part of $\sqrt{n}$. The algorithm uses the continued fraction expansion of $\sqrt{n}$.

Let $x_0 = \sqrt{n}$, and for $i > 0$, $x_i = a_i + 1/x_{i+1}$, where a_i is a positive integer, and $x_i > 1$. We can always write

$$x_i = \frac{\alpha_i + \sqrt{n}}{\beta_i}. \tag{1}$$

In fact, this is true for $i = 0$ with $\alpha_0 = 0$ and $\beta_0 = 1$. If we compute x_{i+1} we get

$$x_{i+1} = \frac{\beta_i(\alpha_i - a_i\beta_i) - \beta_i\sqrt{n}}{(\alpha_i - a_i\beta_i)^2 - n}. \tag{2}$$

This can obviously be written in the form (1) if the following holds:

$$\alpha_{i+1} = a_i\beta_i - \alpha_i \tag{3}$$

$$\beta_i\beta_{i+1} = n - \alpha_{i+1}^2. \tag{4}$$

Consider now $r_i = g - \alpha_{2i-1}$, $s_i = g - \alpha_{2i}$, $q_i = a_{2i}$, $p_i = a_{2i+1}$, $P_i = \beta_{2i}$ and $Q_i = \beta_{2i-1}$. In Equation (3) separate even and odd cases. We get

$$2g - s_i = P_iq_i + r_{i+1} \tag{5}$$

$$2g - r_{i+1} = Q_{i+1}p_i + s_{i+1}. \tag{6}$$

Note that $x_1 = (g + \sqrt{n})/(n - g^2)$ hence $\alpha_1 = g$, $\beta_1 = n - g^2$. Consider now Equation (4). We have $\beta_i\beta_{i+1} = n - \alpha_{i+1}^2$ and $\beta_i\beta_{i-1} = n - \alpha_i^2$. Consider the difference, and use (3). This gives

$$\beta_{i+1} - \beta_{i-1} = a_i(\alpha_i - \alpha_{i+1}). \tag{7}$$

It is now obvious that the quantities α_i and β_i are integers. Separate even and odd cases. We get

$$Q_{i+1} = Q_i + q_i(r_{i+1} - s_i) \tag{8}$$

$$P_{i+1} = P_i + p_i(s_{i+1} - r_{i+1}). \tag{9}$$

Assume for a moment $0 \leq g - \alpha_i < \beta_{i-1}$, in other words, $0 \leq r_{i+1} < P_i$ and $0 \leq s_{i+1} < Q_{i+1}$. This means that (5) and (6) are Euclidean division. Assume s_i, P_i and Q_i are given. Then q_i and R_{i+1} are deduced from (5), Q_{i+1} from (8), p_i and s_{i+1} from (6) and finally P_{i+1} from (9).

Assume $\beta_i > 0$; let us show $0 \leq g - \alpha_{i+1} < \beta_i$ and $\beta_{i+1} > 0$. Since $x_{i+1} = \beta_i/(\sqrt{n} - \alpha_{i+1})$, the condition $x_{i+1} > 1$ is equivalent to $\beta_i > \sqrt{n} - \alpha_{i+1} > 0$. This gives $0 \leq g\alpha_{i+1} < \beta_i$. Now by (3) $\alpha_{i+1} \geq -\alpha_i \geq -g > -\sqrt{n}$ by induction. This implies $\beta_{i+1} > 0$.

As a consequence, we have $-g \leq \alpha_i \leq g$ and $0 \leq \beta_i \leq 2g$. In other words, this process is periodic, the period being at most $4n$.

We introduce now some other quantities, defined by

$$\gamma_{i+1} = \gamma_{i-1} + a_{i+1}\gamma_i \tag{10}$$

$$\delta_{i+1} = \delta_{i-1} + a_{i+1}\delta_i. \tag{11}$$

where $\gamma_{-1} = 1$, $\gamma_{-2} = 0$, $\delta_0 = 1$, $\delta_{-1} = 0$. One can show that these are the numerator and denominator of the approximants to $\sqrt{n}$. Write $A_i = \gamma_{2i-2}$ and $B_i = \gamma_{2i-1}$. Then (10) is equivalent to

$$A_{i+1} = A_i + q_i B_i \tag{12}$$

$$B_{i+1} = B_i + p_i A_{i+1}. \tag{13}$$

Consider finally

$$X_i = n\delta_i^2 - \gamma_i^2 - (-1)^i\beta_{i+1}$$

$$Y_i = n\delta_i\delta_{i-1} - \gamma_i\gamma_{i-1} + (-1)^i\alpha_{i+1}$$

Compute first Y_{i+1}. Replace δ_{i+1} and γ_{i+1} using Equations (10) and (11). Recognise X_i and Y_i. There are some other terms, that vanish because of Equation (3). We get

$$Y_{i+1} = a_{i+1}X_i + Y_i.$$

Compute then X_{i+1}. Replace δ_{i+1} and recognise X_i and X_{i-1}. We get

$$X_{i+1} = a_{i+1}^2X_i + X_{i-1} + 2na_{i+1}\delta_i\delta_{i-1} - 2a_{i+1}\gamma_i\gamma_{i-1} + (-1)^i\beta_{i+2} - (-1)^i\beta_i + (-1)^i a_{i+1}^2\beta_{i+1}.$$

Using (3) and (7), we have $\beta_{i+2} - \beta_i + a_{i+1}^2\beta_{i+1} = 2a_{i+1}\alpha_{i+1}$. Hence

$$X_{i+1} = a_{i+1}^2X_i + X_{i-1} + 2a_{i+1}Y_i.$$

Since $X_0 = 0$, $X_{-1} = 0$, $Y_0 = 0$, we deduce $X_i = 0$ for all i. This means that the following is true:

$$A_i^2 + Q_i = n\delta_{2i-2}^2$$

$$B_i^2 - P_i = n\delta_{2i-1}^2.$$

In general, we shall use these equations in the form $b^2 = r \bmod n$, and compute b modulo n. However, if p is a prime number that divides r, the equation $b^2 = r + \delta^2 n$ implies that n is a square modulo p.

2 Lenstra

The algorithm is based on so-called "elliptic curves". Assume that K is a field, A, B, C, D are elements of K. We consider the set E of points P with homogeneous coordinates (x, y, z) satisfying

$$Azy^2 = x^3 + Bzx^2 + Cz^2x + Dz^3. \tag{0}$$

In the case $z = 0$, we get $x = 0$, hence $y = 1$. This point will be noted O in the sequel. For all other points, we may assume $z = 1$, and consider the equation

$$Ay^2 = x^3 + Bx^2 + Cx + D. \tag{1}$$

For a point P with coordinates (x, y) we denote by $\overline{P}$ the point with coordinates $(x, -y)$. If P is on the curve, this new point will also be on it. We shall assume that the equation $y = 0$ has only simple roots in x. This will ensure that the tangent at the curve is defined at each point. Given two points P_1 and P_2, let P_1P_2 be the third point of intersection of the line passing through P_1 and P_2 with the curve (the tangent if both points are equal). If this point is Q, then $\overline{Q}$ will be called the sum of the two points and denoted by $P_1 + P_2$.

Assume $P_3 = P_1 + P_2$, and the coordinates of P_i are x_i and y_i. We have then the following equations:

$$\delta = \frac{3x_1^2 + 2Bx_1 + C}{2Ay_1} \text{ if } P_1 = P2 \qquad \delta = \frac{y_1 - y_2}{x_1 - x_2} \text{ otherwise.} \tag{2}$$

$$y_2 + y_3 = \delta(x_2 - x_3) \qquad x_1 + x_2 + x_3 + B = A\delta^2. \tag{3}$$

Proof. As we shall see these formulae are only useful if no point is at infinity. It is in fact clear that $O + P = P + O = P$, and that $P + \overline{P} = O$.

Assume first that the two points are distinct. Write $x_3 = x_2 + \lambda(x_1 - x_2)$ and $-y_3 = y_2 + \lambda(y_1 - y_2)$. Eliminate λ between these two equations. We get $y_2 + y_3 = \delta(x_2 - x_3)$. In the case where the two points are equal, these equations have to be replaced by $x_3 = x_2 + \bar{x}$ and $-y_3 = y_2 + \bar{y}$ where the quotient of $\bar{x}$ and $\bar{y}$ is the slope of the tangent. This gives the same equation.

Consider the difference $Ay_2^2 - Ay_3^2 = A(y_2 - y_3)(y_2 + y_3)$. Write it in function of x_2 and x_3, and introduce δ. This gives

$$A\delta(y_2 - y_3) = x_2^3 + x_2x_3 + x_3^2 + Bx_2 + Bx_3 + C. \tag{$*$}$$

We have the same equation with x_1 instead of x_2. Consider the difference of these two equations and factor out $\delta = (y_2 - y_1)/(x_2 - x_1)$. This gives the second equation when the two points are different. If they are equal, just consider the difference of Equation $(*)$ and $2Ay_2\delta = 3x_2^2 + 2Bx_2 + C$.

Lemma: Consider the projective complex plane. If two cubics, defined by homogeneous polynomials F_1 and F_2, have no component in common, then each homogeneous cubic F that passes through eight of their intersection points passes through the ninth point also.

We consider two points distinct points A and B. We can always chose constants c_1 and c_2 such that the curve defined by

$$F^* = F_{A,B} = F - c_1F1 - c_2F_2$$

passes through A and B. In the case $F^* = 0$, we have finished. Assume hence that it is nonzero, its degree is between 1 and 3. If $P_1, \ldots, P_8$ are the eight points common to the three

cubics, $F_{A,B}$ passes through $P_1, \ldots, P_8$ as well as A and B. Now at most three of $P_1, \ldots, P_8$ can be on a line; otherwise this line will be a common component of $F_1 = 0$ and $F_2 = 0$. Similarly, at most six of these points lie on a conic (this is the Bezout Theorem). Out of $P_1, \ldots, P_8$, two, say P_1 and P_2, always lie on a line L and five, say $P_4, \ldots, P_8$, lie on a conic C. There are three cases to be considered: In the first case, P_3 lies on L. We take a point A on L and B neither on L nor C. Because L and $F^* = 0$ have four points P_1, P_2, P_3 and A in common, L is a component of $F^* = 0$. The other component must be C. This contradicts the fact that B is on $F^* = 0$. In the case where P_3 lies on C, we chose A on C, B not on C neither L. On this case $F^* = 0$ intersects C with at least 5 five points, hence C is a component of F^*. The other component must be L. Finally, if P_3 is neither on C nor L, we chose A and B on L.

We claim that $+$ this defines a commutative group on E. The nontrivial part is to prove associativity. We consider only the general case. We have to prove that $X = Y$ where $X = P_1(P_2 + P_3)$ and $Y = (P_1 + P_2)P_3$. We have to prove a big algebraic identity, depending on parameters A, B, C and D. It suffices to prove this identity if the parameters are complex numbers. Let C_1 be the cubic formed by the three lines L_1, L_2 and L_3, where L_1 passes through P_1, P_2 and P_1P_2, L_2 passes through P_3, $P_1 + P_2$ and $P_3(P_1 + P_2)$ and L_3 passes through P_2P_3, $P_2 + P_3$ and O. All these nine points are on E. Let C_2 be the cubic formed by the three lines l_1, l_2 and l_3, where l_1 passes through P_3, P_2 and P_3P_2, l_2 passes through P_1, $P_2 + P_3$ and $P_1(P_2 + P_3)$ and l_3 passes through P_2P_1, $P_2 + P_1$ and O. These nine points are also on E. It suffices to apply the lemma.

Simpler formulae. Assume that the coordinates of P are x and y, that of $2P$ are x' and y'. Then

$$x' = \frac{(x^2 - C)^2 - 4D(2x + B)}{4(x^3 + Bx^2 + Cx + D)}. \tag{5}$$

The proof is easy: consider $4A^2y^2\delta^2 = (3x^2 + 2Bx + C)^2$. Replace $A\delta^2$ by (3), Ay^2 by (1) and expand. Assume now that we want to compute $P + Q$ knowing $P - Q$. If x_0, x_3 are the x-coordinates of $P \pm Q$ and x_1, x_2 are the coordinates of P and Q, then

$$x_0x_3 = \frac{(x_1x_2 - C)^2 - 2(B + x_1 + x_2)D}{(x_1 - x_2)^2}. \tag{6}$$

The proof is easy: consider $\delta = (y_1 - y_2)/(x_1 - x_2)$ and $\delta' = (y_1 + y_2)/(x_1 - x_2)$. We know that $x_1 + x_2 + x_3 + B = A\delta^2$, $x_1 + x_2 + x_0 + B = A\delta'^2$. Compute x_0x_3 from these two equations. The quantities $\delta^2 + \delta'^2$ and $\delta^2\delta'^2$, hence $y_1^2 + y_2^2$ and $(y_1^2 - y_2^2)^2$ are needed. They are easily obtained from (1). Expand.

Given a point P and an integer N, we want to compute NP. We assume here that N is prime. In general, to compute this, the well know binary method is used. In order to use Equations (5) and (6), the binary expansion of N has to be precomputed, and at each stage, mP and $(m + 1)P$ is computed. We need either $2mP$ or $(2m + 2)P$, this is a duplication, and $(2m+1)P$, this is the sum, knowing that the difference is P. Our algorithm is sometimes faster (should compute the mean complexity). Let N_0 be N. Consider three integers a, b and c such that $\pm c = a - b$, $aN + bM = N_0$. We assume $A = aP$, $B = bP$ and $C = cP$. The value of M is arbitrary, we assume however $0 \leq M < N$. We begin with $a = 1$ and $b = 0$. Each time the equation is replaced by $(a + b)N + b(M - N) = N_0$ or $a(N - M) + (a + b)M = N_0$. We modify the biggest of N and M. This means that the max is strictly decreasing. Since N_0 is prime, N and M are always coprime. This means that if $N = M$ or $M = 0$ then $N = 1$. The algorithm stops when $M = 0$, hence $a = N_0$. The first iteration is special: because $M < N$, we replace b with $a + b$, that is, with a. The second iteration is also special: we have to compute $a + b = 2$, hence $2P$. To minimise to number of operations, the ratio N/M should be the Golden Ratio.

The next phase will be to change the value of the prime number. We need however to explain a bit more what happens. Let p be a prime number, F_p the field of integers modulo p. On this field, we define a group structure. The order L (number of elements) is easy to define. For each i let y_i be the quantity $(i^3 + Bi^2 + Ci + D)/A$. If this is 0, there is one point with i as x-coordinate. If this is a square, there are two points, otherwise none. Recall that $\left(\frac{a}{p}\right)$ is the Legendre symbol, it is 0 if a is 0, 1 if a is a square modulo p, -1 otherwise. If we do not forget the point at infinity, the order will be

$$L = 1 + \sum_{i=0}^{p-1} [\left(\frac{y_i}{p}\right) + 1]. \tag{7}$$

This cannot be used easily. One can show that

$$p + 1 - 2\sqrt{p} \leq L \leq p + 1 + 2\sqrt{p}. \tag{8}$$

Any element of the interval can be an order, and the orders are fairly well distributed in the interval. In fact, we shall chose the curve randomly. The idea is now to compute LP for a point P. This will be the point at infinity. The denominator of the x-coordinate will be 0 modulo p. Assume that p is a factor of n. If we are lucky, it will not be zero modulo n. Of course, we do not compute LP, but multiply P by a power of 2, then 3, then 5, etc. We never compute the y coordinate. We assume that formulae (5) and (6) will give x and z, and if they are both zero mod p, that y is not zero mod p.

Instead of considering a general curve, $Ay^2 = x^3 + Bx^2 + Cx + D$ we can make B or D vanish, make A or C unity. Some people consider $y^2 = x^3 + ax + b$. We prefer $Ay^2 = x^3 + Bx^2 + 1$. The main reason is that Equations (5) and (6) simplify a lot.

Assume $B' = (B + 2)/4$, $x = a/b$ is rational in (6). Then x' is

$$x' = \frac{(a^2 - b^2)^2}{4ab[(a - b)^2 + 4abB']}.$$

This can be computed in one less multiplication:

$$t_1 = (a + b)^2 \qquad t_2 = (a - b)^2 \qquad x' = \frac{t_1 t_2}{(t_1 - t_2)[t_2 + B'(t_1 - t_2)]}$$

In the same fashion, (6) can be simplified. If we assume $x_1 = a/b$ and $x_2 = a'/b'$ we have

$$t_1 = (a - a')(b + b') \qquad t_2 = (a + a')(b - b') \qquad x_0 x_3 = \frac{(t_1 + t_2)^2}{(t_1 - t_2)^2}$$

We shall implement these equations. A point P is represented by two numbers `Px` and `Pz`. We assume that the point A is the initial point. We have to allocate space for points B and C, for the sum $D = A + B$, and for t_1, t_2.

The problem is now to chose a curve and a point on it. We begin with choosing a point Q, with coordinates a and b such that $3Q = 0$. This is easy to construct: Q and $2Q$ must have the same x-coordinates. Using (5) gives

$$B = \frac{-3a^4 - 6a^2 + 1}{4a^3} \qquad A = \frac{(a^2 - 1)^2}{4ab^2}. \tag{11}$$

Assume once and for all n coprime to 6. The value of b is really unimportant. We want a to be coprime to n, so that A and B are well-defined. We also want a^2-1 to be coprime with n, so that A is nonzero.

We now chose an initial point P_0. For instance $x=3a/4$. This is nonzero. In order for y to exist, $Ax(x^2+Bx+1)$ must be a square. If we expand, we find that $9-6a^2$ must be a square. exist, the quantity $9-6a^2$ has to be a square. $a=6r/(r^2+6)$. This gives

$$y=\frac{3b(r^2-6)}{4(r^2+6)(a^2-1)}. \tag{12}$$

The quantity r is randomly chosen.

We now have to check that $a(a^2-1)$ is invertible modulo n. In fact, we add another test. Note that the point P such that $x=\pm 1$ is such that $x'=0$ for $P'=2P$. This means that $4P=0$. The question: is this point on the curve? Consider a prime number p, coprime to $A(B-2)(B+2)$. In the case $B(A+2)$ is a quadratic residue, then $yA=\sqrt{A(B+2)}$ for $x=1$. In the case when $B(A-2)$, just chose $x=-1$. In the case where one and only one of $B\pm 1$ is a residue, then one of $A(B\pm 2)$ is a residue. In the other case, B^2-4 is a residue, say $B^2-4=\Delta^2$. In this case, $x=(-B\pm\Delta)/2$ and $x=0$ are three roots of $y=0$. We do not have $4P=0$, but three points such that $2P=0$. This means that modulo each prime p, the order is always a multiple of 4. The primes under consideration are only divisors of n.

Note that if B^2-4 is not a square, then $y=0$ has a simple root. If the case where this is a square, it has three roots. It is easy to see that this quantity has $16a^6$ as denominator. The numerator is $(a^2-1)^3(9a^2-1)$. To avoid multiple roots, we add the condition that $9a^2-1$ is invertible modulo n.