4. Primes and Composites

We have seen that it is necessary to find large primes quickly in order to generate cryptosystems. The cryptanalyst's job is to factor numbers into prime factors (or at least find many prime factors of a number). We examine these problems in this section.

4.1. **Eratosthenes' sieve.** The most traditional way of finding prime numbers is the sieve of Eratosthenes. Suppose we are given a list l of all primes less than the integer n. Then n is prime if it is not a multiple of any of these. Thus, if we have also a list m, so that m_i is the least multiple of the prime l_i that is not less than n, we can compare n with elements of this list to decide if it is a prime. The incremental step can then be carried out as follows. We run through the list m_i look for n. Whenever we find n we note that n is composite and replace m_i by $m_i + l_i$. If we come to the end of the list without find such an i, then n is a prime so we append it to the end of l; we also append $2 * n$ to the end of m. In this way we can iteratively generate the list of all primes!

Clearly as the list grows larger this is taking up more and more space and time. Moreover, it gives us no way of checking if a given number is prime except by running through all primes before it.

4.2. **Trial division.** If $N = a \cdot b$ is a factoring of a number N, then at least one of the numbers a, b is such that its square is not more than N. Thus to check, whether N is a prime, it is enough to test whether it is a multiple of prime number p so that p^2 is not more than N. This leads to the first test that does not require a list of all primes less than N. Given the increasing sequence l of all primes less than some number x so that $x^2 > N$, we check for the primality of N as follows. We run through the sequence l, dividing N by the primes l_i to obtain $N = q_i l_i + r_i$. If r_i is zero then N is not a prime and we stop. If $q_i < l_i$, then $l_i^2 > N$ so have checked enough prime factors to show that N is a prime and we stop.

4.3. **Combinations of the methods.** When we compare the trial division method with Eratosthenes' sieve, we see that we are only checking for divisors upto $\sqrt{N}$, *but* we are performing divisions rather than the (much simpler) additions. Thus there should be a way of improving the Eratosthenes' method.

First of all, when we find a new prime m we should append the *square* m^2 to the list of multiples rather than $2 * m$; all smaller multiples of m are also multiples of smaller primes! This is not enough to speed up the sieve computation since we will still be comparing our trial prime m with all the elements in the multiple list instead of only the "relevant" ones; the list of multiples grows as fast as the list of primes, whereas the multiples we need to check against are only multiples of primes less than the square root. Thus we keep a pointer s into the list m that keeps track of where the list of squares in this list start. We do not check for multiples beyond this point. This extension of the sieve using the idea from trial division is a good way to generate lists of primes.

Trial division is then only a method to check for primality for a small set of numbers and not a method for building lists. Its one big limitation is its dependence on a list of all primes upto $\sqrt{N}$. We do not want to keep extending our list of primes (otherwise we may as well be generating using the sieve as above). So we can ask if we can improve trial division using ideas from the sieve.

It is possible to append a "sieve" at the end of the given list of primes l as follows. Let M the product of a the first few prime numbers (say 6 or 30 or 210). Let S be the (naturally ordered) collection of representatives between 0 and $M-1$ of the units in $\mathbb{Z}/M\mathbb{Z}$.

After we exhaust the given list l without obtaining $q_i < l_i$ we can "extend" the trial division process by trying divisors of the form $s + nM$ where s runs over the residue classes S and n over non-negative integers. While these divisors are not primes, they include all primes and we are at least eliminating some "obvious" repetitions.

Algorithmically, we can apply the above procedure as follows. Let $s_1 < s_2 < \cdots < s_r$ be the list of elements of S. We place the residue class modulo M of the largest element p of l (which should be larger than the factors of $M!$) in this list—say as s_t. After we have exhausted the list l we try $l_{k+1} = l_k + (s_{t+1} - s_t)$. Then we can try $l_{k+i+1} = l_{k+i} + (s_{i+1} - s_i)$ in succession, with the understanding that the successor s_{r+1} or s_r is $M + s_1$; and more generally $s_{ar+b} = aM + s_b$.

This allows us to check the primality of numbers larger than the square of p as well. However, the job of running through a long list of divisors makes trial division unsuitable for finding large prime factors. One can show that (for many numbers) trial division quickly finds small prime factors and then spends a lot of time running through the lists trying to find the larger ones.

4.4. **Compositeness Tests.** We now look at tests that will try to show that a number is composite. In other words, the test either shows that the number is composite or exits (*apparently*) without giving any information.

If p is a prime number, then all numbers between 1 and $p-1$ give units in $\mathbb{Z}/p\mathbb{Z}$. In fact, $\mathbb{Z}/p\mathbb{Z}$ is a *field* and we have the elementary result

Lemma 6. *The group of units in a finite field is a cyclic group.*

Proof. By Legendre's theorem we see that the order of any element divides the order of the group. On the other hand, if x has order dividing d then it is a solution of $T^d - 1$; the latter has at most d solutions since we are in a field. Thus, the exponent of the group of units (i. e. the least common multiple of the orders) must be equal to the order of the group. Now, given elements x and y of orders m and n in an abelian group it is easy to construct an element of the form $x^a y^b$ which has order equal to the least common multiple of m and n. Thus we have a unit of order equal to the order of the group of units; in words the group is cyclic. □

In particular, by Legendre's theorem we see that $a^{p-1} = 1$ in $\mathbb{Z}/p\mathbb{Z}$ for any non-zero element a. Thus if we wish to check whether a number N is composite we can try to find a so that $a^{N-1} \neq 1$ in $\mathbb{Z}/N\mathbb{Z}$. This is already a good check to see that N does not have square factors.

Lemma 7. *When N is an odd number that has square factors, let us define the set of "bad" elements S*

$$S = \{a \in \mathbb{Z}/N\mathbb{Z} | a^{N-1} = 1\}$$

Then, the cardinality of S is at most $2N/9$. If N has no prime factors smaller than p this can be improved to $(p-1)N/p^2$.

The proof depends on the following very important result

Proposition 8. *The group of units in $\mathbb{Z}/p^e\mathbb{Z}$ is cyclic for any* odd *prime number p and any $e \geq 1$.*

We will defer the proof of this proposition to the next subsection.

Proof. (of the lemma) Since N is odd and has square factors there is an odd prime p and an $e \geq 2$ so that p^e is the exact power that divides N. Any element $a \in S$ gives an element b in $\mathbb{Z}/p^e\mathbb{Z}$ so that $b^{N-1} = 1$. Since the latter group is cyclic of order $p^e - p^{e-1}$ it follows that the number of possible values of b is $\gcd(N-1, p^e - p^{e-1})$ (exercise!). Since p divides N, this GCD is equal to $\gcd(N-1, p-1)$, which is not more than $p-1$. The fraction of such b's is thus at most $(p-1)/p^2$ (since $e \geq 2$). By the Chinese remainder theorem, the set S is the same fraction of elements of $\mathbb{Z}/N\mathbb{Z}$. □

While this result is useful to know, one can write numbers (which are called Carmichael numbers) such as $N = 561 = 3 \times 11 \times 17$, with the property that the order of *every* unit in $\mathbb{Z}/N\mathbb{Z}$ divides $N-1$. The necessary improvement on the test was suggested by Miller and Rabin.

We write $N = 1 + q2^k$ with q odd. Now, when N is a prime, $\mathbb{Z}/N\mathbb{Z}$ is a field. Thus, the only element other than 1 whose square is 1 is -1. It follows that for any $a \neq 0$, either $a^q = 1$ or there is some e between 0 and $k-1$ so that $a^{q2^e} = -1$. Now we have seen that computing powers in $\mathbb{Z}/N\mathbb{Z}$ is easily done. Thus we can pick any a and form the powers a^{q2^e} for $0 \leq e < k$ in succession. If $a^q \neq 1$ and none of these powers is -1, then N must be composite. On the other hand, it could happen that for all the a's we pick either $a^q = 1$ or some $a^{q2^e} = -1$. In this case we appear to have obtained no information. However, we have

Lemma 9. *Let N be a composite number of the form $1 + q2^k$. Let us define the set of "bad" elements*

$$T = \{a \in \mathbb{Z}/N\mathbb{Z} | a^q = 1 \text{ or } a^{q2^e} = -1 \text{ for some } e \text{ with } 0 \leq e < k \}$$

Then, the cardinality of T is less than $N/4$.

Proof. Let us write the prime factorisation $N = p_1^{e_1} \cdots p_r^{e_r}$. Now, if a is in T, then clearly $a^{q2^k} = 1$, so a is also in the set S defined earlier. Since $2N/9 < N/4$ (!) we may as well assume that $e_i = 1$ for all i. In other words, we assume that N is a product of distinct prime factors. Now, we write $p_i = 1 + q_i 2^{k_i}$ with q_i odd; for later use we note that k is not less than the minimum of the k_i's. We further decompose T into the set $T_{-1} = \{a | a^q = 1\}$ and the sets (for $0 \leq e < k$)

$$T_e = \{a | a^{q2^e} = -1\}$$

Then, elements of T_{-1} reduce to units in $\mathbb{Z}/p_i\mathbb{Z}$ which have order dividing q. This is a subgroup of order $\gcd(q, p_i - 1) = \gcd(q, q_i)$. Thus, by the Chinese remainder theorem

$$\#T_{-1} = \gcd(q, q_1) \cdots \gcd(q, q_r)$$

The elements of T_e, can be characterised as elements, whose q-th power has order *exactly* 2^{e+1}. These q-th powers then have order exactly 2^{e+1} when reduced modulo p_i. In particular, this means that $e < k_i$ for every i; the other T_e's are empty. There are exactly $\gcd(q, q_i)2^{e+1}$ elements in $\mathbb{Z}/p_i\mathbb{Z}$ with order dividing $q2^{e+1}$ and among

these a subgroup of index 2 has elements of order dividing $q2^e$ (a subgroup of a cyclic group is cyclic). Thus, by Chinese remainder theorem we obtain

$$\#T_e = \gcd(q, q_1) \cdots \gcd(q, q_r) \cdot 2^{re}$$

Thus we see that the cardinality of T is

$$\gcd(q, q_1) \cdots \gcd(q, q_r) \left(1 + \sum_{i=0}^{l-1} 2^{re}\right) = \gcd(q, q_1) \cdots \gcd(q, q_r) \left(\frac{2^{rl} + 2^r - 2}{2^r - 1}\right)$$

where l is the minimum of the k_i's. Now, the Chinese remainder theorem shows that the number of units in $\mathbb{Z}/N\mathbb{Z}$ is precisely $q_1 \cdots q_r \cdot 2^{\sum_i k_i}$; this is at least one less than N. Thus the proportion of elements in T is strictly smaller than

$$\frac{\gcd(q, q_1) \cdots \gcd(q, q_r)}{q_1 \cdots q_r} \cdot \frac{2^{rl} + 2^r - 2}{(2^r - 1) \cdot 2^{k_1 + \cdots + k_r}}$$

The first term is no more than 1, while the second is no more than $1/2^{r-1}$ (note that $l \geq 1$). Thus, we obtain the result unless $r = 2$. Moreover, if $k_2 > k_1$ (or vice versa) then we see that the second term is no more than $1/2^r$ so we have the result in this case as well. Thus we may assume that $k_1 = k_2 = l$. Now, if $\gcd(q, q_1) < q_1$ then (since these are odd numbers and one divides the other) $\gcd(q, q_1) \leq 3q_1$. This implies that the above expression is no more than $1/6$. Thus, we may further assume that $\gcd(q, q_1) = q_1$. By expanding the identity $(1 + q2^k) = (1 + q_1 2^l)(1 + q_2 2^l)$ we see that $\gcd(q, q_1) = \gcd(q, q_2)$. Since the primes p_1 and p_2 are distinct $q_1 \neq q_2$; thus $q_1 = \gcd(q, q_1) \leq 3q_1$ as above. Now we again obtain that the above expression is no more than $1/6$. This completes the argument. □

What the above reasoning amounts to is that if we choose *uniformly* among all possible a's in $\mathbb{Z}/N\mathbb{Z}$, there is a chance of less than $1/4$ that we will pick an a which gives "no information" as the output of our test even though N is composite. This is not "no information" at all! If we repeat this test n times there is a chance of less than $(1/4)^n$ that N is composite and we did not detect it. It seems more than reasonable to call an N that satisfies such a test a *strong pseudo-prime*. When we specify the a_1, a_2, …, a_n, we say that N is a strong pseudo-prime with *bases* a_1, a_2, …, a_n.

While have not actually proved that N is a prime in such a case (unlike trial division) there appears to be good enough reason to treat it like a prime. In later sections we will look at primality tests and primality certificates. It is clear that we should not even attempt those unless we have already put our N through the Miller-Rabin grinder and it has come out successful!

4.5. **Hensel's lemma.** We conclude this section with a proof of the proposition 8. The group of units in $\mathbb{Z}/p^e\mathbb{Z}$ is of order $p^e - p^{e-1} = p^{e-1}(p-1)$, thus it is enough to find elements of order $p-1$ and p^{e-1} in this group. First we prove a result that will be useful in other contexts

Lemma 10 (Hensel's lemma)**.** *Let $f(T) = T^d + a_1 T^{d-1} + \cdots + a_d$ be a (monic) polynomial with integer coefficients a_i. Let n be an integer so that $f(n)$ is divisible by p, and $f'(n) = dn^{d-1} + a_1(d-1)n^{d-2} + \cdots + a_{d-1}$ is not divisible by p. Then there is a sequence of integers n_k for every $k \geq 1$, so that $n_1 = n$, $n_{k+1} - n_k$ is divisible by p^k and $f(n_k)$ is divisible by p^k. Moreover, n_k is uniquely determined modulo p^k.*

Proof. The proof closely mimics Newton's method of finding roots. Having already found n_k we need to find $n_{k+1} = n_k + b_k p^k$ so that $f(n_{k+1})$ is divisible by p^{k+1}. By the binomial expansion (or Taylor series!),

$$f(n_k + b_k p^k) = f(n_k) + b_k p^k f'(n_k) \pmod{p^{k+1}}$$

using $2k \geq k+1$ since $k \geq 1$. We are given $f(n_k) = c_k p^k$ for some constant c_k. Moreover, $n_k = n \pmod{p}$ so $f'(n_k) = f'(n) \pmod{p}$ is an invertible element of $\mathbb{Z}/p\mathbb{Z}$. Let m be an inverse so that $mf'(n_k) = 1 \pmod{p}$. We put $b_k = -mc_k$ and obtain the required condition. □

We now apply this to the polynomial $f(T) = T^{p-1} - 1$ and any integer n prime to p to conclude that there is an integer n_e so that $n_e^{p-1} = 1 \pmod{p^e}$ (note that $f'(T) = -T^{p-2} \pmod{p}$). This gives us the required elements of order $p-1$ (since such exist modulo p). Moreover, we see that the units in $\mathbb{Z}/p^e\mathbb{Z}$ can be written as $g^a u$ where g is an element of order $(p-1)$ and $u = 1 \pmod{p^e}$. Let U_1 be the group of elements of the latter kind. We will now apply log and exp in a suitable way to conclude the result.

Lemma 11. (1) *Let x be divisible by p then the power of p that divides x^n/n is at least $n - [\log_p(n)]$, where the latter term denotes the integral part of $\log_p(n)$. In particular, this goes to infinity with n.*

(2) *Let x be divisible by p, then the power of p that divides $p^n/n!$ is at least $n - \sum_{i>0}[n/p^i]$. In particular, if p is odd then the latter term goes to infinity with n.*

Proof. Exercise. □

It follows that only finitely many terms of the power series

$$\log(1-x) = -\sum_{i \geq 0} \frac{x^i}{i+1}$$

survive in $\mathbb{Z}/p^e\mathbb{Z}$ when we substitute x by a multiple of p. Thus we obtain a map

$$\log : U_1 \to \mathbb{Z}/p^e\mathbb{Z}$$

which in fact takes values in the ideal $p\mathbb{Z}/p^e\mathbb{Z}$, which isomorphic to the additive group $\mathbb{Z}/p^{e-1}\mathbb{Z}$. Elementary manipulations of the power series combined with the binomial theorem and the fact that all but finitely many terms are zero can be used to show that $\log(1 + x \cdot y + x + y) = \log(1+x) + \log(1+y)$. Similarly, for p odd we obtain a map

$$\exp : p\mathbb{Z}/p^e\mathbb{Z} \to U_1$$

by means of the usual power series

$$\exp(x) = \sum_{i \geq 0} \frac{x^i}{i!}$$

which satisfies $\exp(x+y) = \exp(x) \cdot \exp(y)$. We also check by direct substitution that $\log(\exp(x)) = x$ and vice versa. It follows that the group U_1 is isomorphic to the group $p\mathbb{Z}/p^e\mathbb{Z}$ which is in turn isomorphic to $\mathbb{Z}/p^{e-1}\mathbb{Z}$. We note in passing that a generator g of U_1 of order p^{e-1} corresponds to the generator p in $p\mathbb{Z}/p^e\mathbb{Z}$ via the expression $g = \exp(p)$!