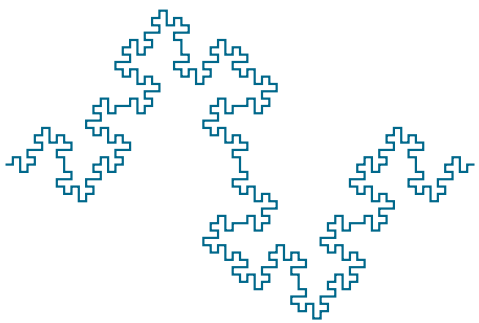
Cryptographic Mathematics I
Week 4

Dr. Eberhard Mayerhofer

University of Limerick



Semester I 2016/7

Euclid's Theorem

Theorem

There are infinitely many prime numbers.

Proof.

(indirect argument) If there were only finitely many primes, we could list them all

$$2 = p_1 < p_2 < \cdots < p_N.$$

Consider $m = p_1 p_2 \cdots p_N + 1$ which is a product of primes by the Fundamental Theorem of Arithmetic. None of the primes $p_1, p_2, \ldots, p_N$ divides m (why?). Thus m must be a prime, but m is bigger than p_N the largest prime according to our assumption. This is a contradiction. □

There is one even prime 2 and all others are odd. Some of these odd primes are congruent to 1 mod 4 and the others to 3 mod 4.

5	13	17	29	37	41	53	61	73	$\ldots \equiv 1 \mod 4$
3	7	11	19	23	31	43	47	59	$\ldots \equiv 3 \mod 4$

In fact there are infinitely many of each of these two types.

Euclid's result extended...

We can prove that there are an infinite number of primes $\equiv 3 \mod 4$ by using Euclid's argument but the same argument does not work for primes that are congruent to $1 \mod 4$.

Proof...?

Hint: Two primes p_1, p_2 which are $\equiv 3$ modulo 4 satisfy

$$p_1 p_2 \equiv 1 \mod 4.$$

Consider then $p_1 p_2 \ldots p_n + 2$ modulo n and distinguish the cases n even and n odd.

Looking at remainders on division by 5 we find $5 \equiv 0 \mod 5$ and a partition of the remaining primes into 4 subsets.

11	31	41	61	71			$\equiv 1 \mod 5$
2	7	17	37	47	67	97	$\equiv 2 \mod 5$
3	13	23	43	53	73	83	$\equiv 3 \mod 5$
19	29	59	79	89			$\equiv 4 \mod 5$

Dirichlet's Theorem

Theorem

If $\gcd(a, m) = 1$ *then there are an infinite number of primes congruent to* $a \mod m$.

The proof of this uses advanced calculus (complex analysis)...

Let $\pi(x)$ be the number of primes that do not exceed x.

Theorem

$\pi(x)$ is approximately equal to $\frac{x}{\ln(x)}$.

Here $\ln(x)$ is the natural logarithm of x or the logarithm to the base e. The number e is approximately 2.7182818.
From this formula we get, for example:

- One quarter of the first 100 positive integers are prime.
- About 2% of numbers less than 10^{22} are prime and 1% of numbers less than 10^{44} are prime.
- in other words: The fraction of numbers being prime decreases, but slowly.

A recent major result about prime numbers by **Green and Tao** has shown that there are arbitrarily long sequences of primes in arithmetic progression.
However, there are many simply stated **conjectures** about prime numbers which have not been proved:

- Goldbach's conjecture: every even number $n \geq 4$ is a sum of two primes. There has been recent progress on this problem but it is not yet solved.
- Twin Primes conjecture: There are an infinite number of primes p for which $p + 2$ is also prime.
- $N^2 + 1$ conjecture: There are infinitely many primes of the form $N^2 + 1$.
- Mersenne Primes conjecture: There are an infinite number of primes p for which $2^p - 1$ is also prime.
- The most important unsolved problem in mathematics is the Riemann Hypothesis. A solution of this will give us a better understanding of the distribution of primes.

ROWLAND'S FORMULA

The search for a formula for the nth prime number is another unsolved problem. However, a remarkable result was recently discovered by Eric Rowland. Here is **Rowland's formula**. We define $a_1 = 7$, and for $n \geq 2$ we set

$$a_n = a_{n-1} + \gcd(n, a_{n-1}).$$

So, for example, we find $a_2 = a_1 + \gcd(2, 7) = 8$. The prime generator is then $a_n - a_{n-1}$, the first differences of the original sequence.

For example, here are the first 23 values of the sequence $a_1, a_2, a_3, \ldots$:

7, 8, 9, 10, 15, 18, 19, 20, 21, 22, 33, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 6

and here are the first differences of these values:

$$1, 1, 1, 5, 3, 1, 1, 1, 1, 11, 3, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 23.$$

If we ignore the 1's, then, the Rowland formula starts by generating the primes 5, 3, 11, 3 (again), and 23. Removing duplicates, the first few are

5, 3, 11, 23, 47, 101, 7, 13, 233, 467, 941, 1889, 3779, 7559, 15131, 53, 30323

Why does it work? The proof is too involved to give here, but it is not that difficult. See Rowland's paper *A Natural Prime-generating Recurrence* (Journal of Integer Sequences Vol 11 (2008)) for the details.

Calculating Large Powers

- Recall: Can use Fermat's Little Theorem to verify that a given large integer n is not a prime number by finding an integer a for which

$$a^{n-1} \not\equiv 1 \mod n.$$

- The advantage of this method is that we do not need to factorise n.
- The disadvantage is that we have no conclusive answer if it happens that $a^{n-1} \equiv 1 \mod n$.
- In order to use this method we need to be able to calculate very high powers of integers $\mod n$. We use: **successive squaring**.

EXAMPLE

Suppose we want to calculate $5^{843} \mod 613$. We carry out the following calculations in which each new number is obtained by squaring the previous.

$$5^1 \equiv 5 \equiv 5 \mod 613$$

$$5^2 \equiv 25 \equiv 25 \mod 613$$

$$5^4 \equiv 25^2 \equiv 12 \mod 613$$

$$5^8 \equiv 12^2 \equiv 144 \mod 613$$

$$5^{16} \equiv 144^2 \equiv 507 \mod 613$$

$$5^{32} \equiv 507^2 \equiv 202 \mod 613$$

$$5^{64} \equiv 202^2 \equiv 346 \mod 613$$

$$5^{128} \equiv 346^2 \equiv 181 \mod 613$$

$$5^{256} \equiv 181^2 \equiv 272 \mod 613$$

$$5^{512} \equiv 272^2 \equiv 424 \mod 613$$

Now $843 = 512 + 256 + 64 + 8 + 2 + 1$. Thus

$$\begin{aligned} 5^{843} &\equiv 5^{512+256+64+8+2+1} \\ &\equiv 5^{512} \cdot 5^{256} \cdot 5^{64} \cdot 5^{8} \cdot 5^{2} \cdot 5^{1} \\ &\equiv 424 \cdot 272 \cdot 346 \cdot 144 \cdot 25 \cdot 5 \\ &\equiv 23 \mod 613. \end{aligned}$$

In essence we have expressed 843 as a binary number and taken those powers which correspond to a 1 in the binary expansion of 843.

Shorter with Fermat

An initial application of Fermat's Little theorem would have shown that

$$\begin{aligned} 5^{843} &\equiv 5^{612} \cdot 5^{231} \\ &\equiv 5^{231} \\ &\equiv 5^{128} \cdot 5^{64} \cdot 5^{32} \cdot 5^{4} \cdot 5^{2} \cdot 5^{1} \\ &\equiv 181 \cdot 346 \cdot 202 \cdot 12 \cdot 25 \cdot 5 \\ &\equiv 23 \mod 613. \end{aligned}$$

Why might one want to do such calculations? One can use such computations to encode and decode messages. The ciphertexts created in this way are unbreakable by current methods.

Calculating roots

We know that we can solve the congruence

$$x^k \equiv a \mod m$$

by successively trying $x = 1, x = 2, x = 3, \ldots$. However, this involves too much calculation if m is large. Knowledge of $\varphi(m)$ can help in the calculations.

Example

Solve the congruence

$$x^{179} \equiv 237 \mod 989.$$

As a first step we calculate $\varphi(989)$. Using that $989 = 23 \cdot 43$, we obtain $\varphi(989) = \varphi(23 \cdot 43) = \varphi(23)\varphi(43) = 22 \cdot 42 = 924$. The next step consists in solving

$$ku - \varphi(m)v = 1 \quad \text{i.e.} \quad 179u - 924v = 1.$$

This is possible as $\gcd(179, 924) = 1$. The Euclidean algorithm gives

$$\begin{aligned} 924 - 5 \cdot 179 &= 29 \\ 179 - 6 \cdot 29 &= 5 \\ 29 - 5 \cdot 5 &= 4 \\ 5 - 1 \cdot 4 &= 1 \\ 4 - 4 \cdot 1 &= 0. \end{aligned}$$

From these calculations we obtain $191 \cdot 179 - 37 \cdot 924 = 1$ and therefore

$$\left(x^{179}\right)^{191} = x^{1+37\cdot 924} = x \cdot \left(x^{924}\right)^{37}.$$

As $\varphi(989) = 924$, Euler's formula tells us that $x^{924} \equiv 1 \mod 989$. Thus

$$x \equiv \left(x^{179}\right)^{191} \equiv 237^{191} \mod 989.$$

In the final step we use successive squaring to calculate 237^{191} mod 989.

$$
\begin{aligned}
237^1 &\equiv 237 \equiv 237 \mod 989 \\
237^2 &\equiv 237^2 \equiv 785 \mod 989 \\
237^4 &\equiv 785^2 \equiv 78 \mod 989 \\
237^8 &\equiv 78^2 \equiv 150 \mod 989 \\
237^{16} &\equiv 150^2 \equiv 742 \mod 989 \\
237^{32} &\equiv 742^2 \equiv 680 \mod 989 \\
237^{64} &\equiv 680^2 \equiv 537 \mod 989 \\
237^{128} &\equiv 537^2 \equiv 570 \mod 989
\end{aligned}
$$

$$\begin{aligned} 237^{191} &\equiv 237^{128+32+16+8+4+2+1} \\ &\equiv 570 \cdot 680 \cdot 742 \cdot 150 \cdot 78 \cdot 785 \cdot 237 \\ &\equiv 290 \mod 989 \end{aligned}$$

Remark

The first step in the calculations is to find $\varphi(m)$. This might cause problems. It is easy if we know the factorisation of m into primes. If $m = pq$ where p and q are primes with 100 or more digits, finding this factorisation may be virtually impossible. It is the difficulty in calculating $\varphi(m)$ which makes the cryptographic method work.

Part I of RSA

The two processes described in the last two sections are the basis of the RSA cryptosystem. The first step is encoding the message to be sent into a string of digits using a block code. One possibility is to use the following encoding for letters:

$$A = 10, B = 11, C = 12, \ldots, Z = 35, \text{␣} = 36.$$

Using this the string 'GOOD MORNING' would become

$$162424133622242723182316$$

Part II of RSA

The next step is to choose two large prime numbers p and q and form the product $m = pq$. It should be hard to factorise m. Thus it does not do to choose p and q with $|p - q|$ small as the factorisation of m can be found easily by the difference of squares method described in a later section. We know the value of $\varphi(m) = (p-1)(q-1)$. Now choose k coprime with $\varphi(m)$. We publish m and k. The pair (m, k) is known as our **public key**.

Someone wishing to send us a message converts their message into a string of digits as above, breaks the string of digits into blocks of length less than the length of m. Suppose the the message is now $a_1, a_2, \ldots, a_n$.
Next they calculate

$$\begin{aligned} b_1 &\equiv a_1^k \mod m \\ b_2 &\equiv a_2^k \mod m \\ b_3 &\equiv a_3^k \mod m \\ &\vdots \\ b_n &\equiv a_n^k \mod m \end{aligned}$$

with $0 \leq b_i < m$. The message to be sent is then $b_1, b_2, \ldots, b_n$. They were able to do this because of their knowledge of k and m.

As the receiver, in order to recover the original message, we need to solve the congruences

$$x^k \equiv b_i \mod m$$

and this we know we can do because we know the value of $\varphi(m)$. An outsider intercepting the message does not know $\varphi(m)$ and will have to try to factorise m. By choosing p and q big enough we can make that task virtually impossible.

History

The original idea of a public key cryptosystem was put forward by Diffie and Hellman in 1976. The implementation described above was invented by Ron Rivest, Adi Shamir and Leonard Adleman and gave rise to the name **RSA public key cryptosystem**.

EXAMPLE

Using the example above and choosing $p = 4973$ and $q = 5237$, we have $m = 26043601$ and $\varphi(m) = 26033392$. Take $k = 12123$. Splitting the message 162424133622242723182316 into four blocks of size 7, starting at the right end, we obtain $a_1 = 162$ $a_2 = 4241336$ $a_3 = 2224272$ and $a_4 = 3182316$. Calculation with `sage` produces

$$\begin{aligned} 162^{12123} &\equiv 9612379 \mod 26043601 \\ 4241336^{12123} &\equiv 9617065 \mod 26043601 \\ 2224272^{12123} &\equiv 20289124 \mod 26043601 \\ 3182316^{12123} &\equiv 19319102 \mod 26043601 \end{aligned}$$

Thus, the encoded message is

$$09612379096170652028912419319102$$

Note that, because m has 8 digits, each block is encoded as a block of length 8. That explains the extra zeros in the final message. As m is of length 8, the receiver knows to divide the received message into blocks of size 8 and then remove any leading zeros.