EINFÜHRUNG IN DIE KRYPTOGRAPHIE
INTRODUCTION TO CRYPTOGRAPHY

SOMMERSCHULE FASZINATION MATHEMATIK, 15.06.- 19.06.2018

MICHAEL AND RÓISÍN NEURURER

1. Modular Arithmetic

Adding around the clock.
We can all read a clock and know that if it is 11 o'clock now, then it is 1 o'clock in 2 hours time. In 12 hours it will be 11 o'clock again. So can we say that $11 + 2 = 1$ (on a clock)?

Try the following and think about how you work them out:

(1) If it is 9 o'clock now, what time is it in 57 hours?

(2) If it is Saturday today, what day is it in 11 days?

(3) If it is June now, what month will it be in 1000 months?

(4) If it is midday now, will it be light or dark in 539 hours?

Modulo.
Looking at the clock examples, in mathematical notation we would write:

$$11 + 2 \equiv 1 \pmod{12} \quad \text{and} \quad 12 \equiv 0 \pmod{12} \quad \text{and} \quad 539 \equiv 11 \pmod{12}$$

Look at the following addition tables for mod 5 and mod 6. Make sure you understand them and then have a go at the following exercises.

Addition modulo 5

+	0	1	2	3	4
0	0	1	2	3	4
1	1	2	3	4	0
2	2	3	4	0	1
3	3	4	0	1	2
4	4	0	1	2	3

Addition modulo 6

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Exercises.

(1) Construct the addition table for mod 7.
(2) What is $3^5 \bmod 7$?
(3) Find the x between 0 and 11 which solves $10 + x \equiv 2 \pmod{12}$.
(4) What is $3^{1000} \bmod 5$?

Multiplication mod n.
As well as adding, we can multiply. Have a look at the multiplication tables for $n = 5$ and $n = 6$.

Multiplication modulo 5

·	0	1	2	3	4
0	0	0	0	0	0
1	0	1	2	3	4
2	0	2	4	1	3
3	0	3	1	4	2
4	0	4	3	2	1

Multiplication modulo 6

·	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

There are some key differences. In the table for mod 5 there is a 1 in every row. This means that for every non-zero number a, there is another number b which gives $a \cdot b \equiv 1$. We call such a number an **inverse**. Another difference is that in the table for mod 6 there are "extra zeros": it can happen that two numbers which are not 0 multiply together to give 0. These differences are related. The rows in the mod 6 table which have extra zeros don't have a 1. The difference is due to the fact that 5 is a prime and 6 is not.

Definition 1. A **prime number** is a whole number which is only divisible by 1 and itself.

We will later show that when multiplying mod p, where p is prime, every non-zero integer has an inverse. For now we can show the following:

Lemma 1. If n is not a prime, then there are non-zero integers that do not have inverses modulo n.

Proof. If n is not a prime then there exist $a, b \neq 1$ such that $n = a \cdot b$. Now $a \not\equiv 0 \pmod{n}$ and $b \not\equiv 0 \pmod{n}$, but $n = a \cdot b \equiv 0 \pmod{n}$. Suppose a has an inverse mod n. This means there exists an x such that $a \cdot x \equiv 1 \pmod{n}$. Then, multiplying both sides by b gives $b \cdot a \cdot x \equiv b \pmod{n}$, which gives $0 \equiv b \pmod{n}$ which is not true. Therefore we cannot have such a number x and so a has no inverse mod n. □

2. Euclidean Division and the Euclidean Algorithm

Definition 2. Given two integers a and b with $b \neq 0$ we say b divides a and write $b \mid a$ if there exist an integer q such that $a = b \cdot q$. We say b is a **divisor** of a.

Euclidean division.

Euclidean division (Division mit Rest) is the process of dividing two integers to give a quotient and a remainder. For example. We can divide 25 by 3. This gives us a quotient of 8 with a remainder of 1. In this example 3 is the **divisor**.

Theorem 2. Given two integers a and b, with $b \neq 0$, there exist unique integers q and r such that $a = bq + r$ and $0 \leq r < b$.

Proof. We can prove this geometrically. Suppose $a \geq b$. Take a line of length a. We can divide this up into lines of length b. Either there will be nothing left, and $r = 0$ or there is

a certain length left. This is necessarily less than b, or else we could take another length b out.
Suppose $a < b$. Then we can't take any lines of length b out and so $q = 0$ and $r = a < b$ $\square$

Every pair of integers will always have 1 as a common divisor, but they may have other common divisors. For example, 12 and 30 have 1, 2, 3, and 6 as common divisors. We are interested in the greatest common divisor. In this example, it is 6 as 6 is the largest divisor of both 12 and 30.

Definition 3. Given two integers a and b we say and integer d is the **greatest common divisor** of a and b if:

(1) $d \mid a$ and $d \mid b$, and
(2) If $d' \mid a$ and $d' \mid b$ then $d' \mid d$

We write gcd(a, b)=d.

Definition 4. Two integers a and b are **coprime** if their greatest common divisor is 1. That is, gcd(a,b)=1.

Excersises.

(1) Why does the above definition give the greatest common divisor?
(2) Why can we be sure that there will always be a gcd?
(3) Find gcd(32, 54)
(4) Find gcd(925, 65)

Definition 5. Two integers a and b are **coprime** if their greatest common divisor is 1. That is, gcd(a,b)=1.

Euclidean Algorithm.

The **Euclidean algorithm** is an efficient method for computing the gcd of two integers. If we are trying to compute the gcd of two integers a and b, we can replace the larger of the two numbers by its remainder when divided by the smaller of the two. We repeat this process until we reach a zero remainder.

Example 1. Let's take $a = 1071$ and $b = 462$. We are looking to compute gcd(1071, 462). We divide 1071 by 462 to get a quotient of 2 and a remainder of 147

$$1071 = 2 \times 462 + 147 \qquad q_0 = 2 \qquad r_0 = 147$$

Now we repeat the process with 462 and 147 to get

$$462 = 3 \times 147 + 21 \qquad q_1 = 3 \qquad r_1 = 21$$

.
Repeating the process for a second time gives

$$147 = 7 \times 21 + 0 \qquad q_2 = 7 \qquad r_2 = 0$$

.
We have a zero remainder so the algorithm ends and the last non-zero remainder gives us the gcd of a and b. So we have that gcd(1071, 462)=21.

Exercises.

(1) Compute gcd(925, 65).
(2) Comupte gcd(5671, 342)

3. Bezout's Identity, Fermat's Little Theorem...

In this section we will use modular arithmetic and the Euclidean algorithm to prove some number theory results which we will need on our journey towards cryptography.

Bezout's Identity.

Lemma 3. For coprime integers a, b, there are integers x and y with $ax - by = 1$.

Proof. Let's apply the Euclidean algorithm to the integers a and b, continuing until we have a zero remainder. Once we have stopped we have

$$a = q_0 b + r_0 b = q_1 r_0 + r_1 ... r_{n+1} = q_n r_n + 0$$

Since a and b are coprime we know that $r_n = 1$. Rearranging the above equations we get

$$r_0 = a - q_0 b r_1 = b - q_1 r_0 = b - q_1 a + q_1 q_0 b = a(-q_1) - b(-1 - q_1 q_0) .... 1 = r_n = r_{n-1} - q_n r_{n-2} = ...$$

which will give some required x and y such that $ax - by = 1$.

□

Bezout's identity is in fact a particular case of the fact that we can write the gcd of a and b as a linear combination of a and b. That is, if gcd(a,b)= d. Then there exist integers x and y such that $d = ax + by$. The method for finding x and y described in the proof above is known as the **extended Euclidean algorithm**

Exercise. Find the gcd of 148 and 272 and write it as a linear combination of these integers.

We can now prove the following two statements from earlier in the course.

Lemma 4. Let p be prime. If p divides $a \cdot b$ then p divides a or p divides b.

Excersise. Try and prove the above lemma yourself. Here are some hints if you are stuck:

(1) Consider the following two cases separately: if p divides a, if p does not divide a.
(2) If p does not divide a what does this say about a and p?
(3) Can you apply Bezout's identity?
(4) What can you say mod p?

Lemma 5. Let p be prime. Every non-zero number mod p has an inverse.

Proof. Let a be non-zero mod p. Then a is coprime to p, so by Bezout we have integers x and y such that $ax - py = 1$. This means that $ax \equiv 1 \pmod{p}$ and so x is an inverse of a mod p. □

Question. Why can we assume a coprime to p in the above proof?

We can now use all we have proved so far to prove Fermat's Little Theorem.

Theorem 6 (Fermat's Little Theorem)**.** For any a coprime to p we have $a^{p-1} \equiv 1 \pmod{p}$.

Proof. We look at the numbers a, $2a$, $3a$, ...$(p-1)a$ modulo p. Since a and p are coprime, and the numbers $1, 2, ..., p-1$ are all less than p, none of these numbers are divisible by p, so none of them is 0 mod p.
Can two of these numbers be the same modulo p? Suppose $ak \equiv al \pmod{p}$ with $k \neq l$. Then $a(k-l) \equiv 0 \pmod{p}$ which means p divides $a(k-l)$. But a and p are coprime, so p doesn't divide a and since k and l are less than p, p cannot divide $k-l$. So this cannot happen and the numbers a, $2a$, $3a$, ...$(p-1)a$ are all different modulo p.
It follows that they are the same as the numbers $1, 2, ..., (p-1)$ mod p just in a different order. So we have

$$a \cdot 2a \cdot 3a \cdots (p-1)a = a^{p-1} \cdot 1 \cdot 2 \cdots (p-1) \equiv 1 \cdot 2 \cdots (p-1) \pmod{p}$$

Now, dividing each side by $1 \cdot 2 \cdots (p-1)$ gives us

$$a^{p-1} \equiv 1 \pmod{p}$$

□

4. Cryptography

These days, we have many different ways of communicating with each other. Many of our communications are over internet and mobile networks and there is a possibility that anyone else with access to the internet could gain access to our messages as they are transmitted. Cryptography allows us to keep our communications safe. In fact, people have needed to keep communications secret throughout history: Generals sending instructions to their soldiers; Leaders discussing mutual enemies; revolutionaries plotting their next move. Historically, the people were sending physical paper messages rather than the electronic communications we use today, but the problem of ensuring the message could not be read by anyone other than the intended recipient remained.

Caesar Cipher.

The Caesar cipher is a way of encrypting a message by shifting the alphabet along. For example, Alice wants to send Bob the message

"Run away"

to Bob. They decide in advance that they are going to use a shift of 3. This means each letter gets replaced by the corresponding letter when the alphabet is shifted 3 places.

plaintext	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
ciphertext	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Alice's message becomes:

UXQDZDB

The Caesar Cipher is an example of secret key cryptography also known as **symmetric key cryptography**. It uses a single key (3 in the above example) to both encrypt and decrypt the message.
Mathematically, the Caesar cipher is the addition of a fixed number modulo 26, where we encode A=1, B=2, ... , Z= 26.
In the above example, we first convert Alice's message to a string of integers

18 21 14 1 23 1 25

And then we add 3 modulo 26 to each number to get the encrypted message

21 24 17 4 26 4 2

Exercise.

(1) How does Bob decrypt Alice's message?
(2) Suppose instead of adding modulo 26, we want to multiply each number by a fixed number. If Alice multiplies each number by 13, can Bob decrypt the message?
(3) Determine which numbers are valid keys for the multiplication algorithm and determine how to decrypt the message.

A major problem with symmetric key cryptography is the exchange of the key. Once Alice and Bob have a key they can send encrypted messages between themselves. But how to they exchange the key itself? If they can meet up and talk about it, then why not just tell each other the message in the first place? If Alice sends Bob the key then Carol could intercept the key and subsequently read all their messages. This is the problem solved by **public key cryptography**.

Public key cryptography.
Let's go back to Alice, Bob and their secret message. Alice has a personal padlock to which only she has the key. She can make this padlock publicly available so that anyone can put it on their message, however only she will be able to open it. Bob simply puts one of Alice's padlock on his message and sends it to her. It's safe, because no one can open it but Alice. The padlock is Alice's **public key** while the key to open it is her **private key**. The RSA algorithm is an example of public key cryptography.
Rather than a physical padlock and key, we use what is known as a **trapdoor function**. This is a function that is easy to perform in one direction but extremely difficult to reverse, unless you have some specific information.

Exercises.

(1) Calculate $23 \cdot 37$.
(2) Find the prime factors of 943.

Which of the previous two exercises was easier? Now imagine we had primes with 300 digits in them! RSA is based on this principle that multiplying primes is easy, while factorising numbers into primes is very difficult.

RSA.

- Choose two large distinct primes p, q.
- Compute $n = p \cdot q$ and $\phi = (p-1)(q-1)$.
- Choose an integer e coprime to ϕ with $1 < e < \phi$.

- Find d, the inverse of $e \pmod{\phi}$, i.e. $d \cdot e \equiv 1 \pmod{\phi}$.
- Throw away p, q, ϕ. Keep n, e, d.
- To encrypt a message $x < n$, calculate $y \equiv x^e \pmod{n}$.
- To decrypt y, compute $y^d \pmod{n}$.

The public key is n and e. Anyone can use these numbers to encrypt a message. The private key is d. Only someone who knows d can decrypt the message.
Now we just need to prove that this works! Does $(x^e)^d \pmod{n}$ give back x? First we need the following result:

Lemma 7. Let p, q be different primes. If $x \equiv y \pmod{p}$ and $x \equiv y \pmod{q}$ then $x \equiv y \pmod{p \cdot q}$

Proof. We can write $x = y + kp = y + lq$. This gives $kp = lq$. Since p and q are different primes p divides l and q divides k. So we can write $k = k'q$ and $l = l'p$ which gives $x = y + k'pq = y + l'pq$ which means $x \equiv y \pmod{pq}$. □

Theorem 8. RSA works. That is, for distinct primes p and q, and d, e calculated as above, $(x^e)^d \equiv x \pmod{pq}$.

Proof. By the above lemma, we only need to check mod p and mod q.
If $x \equiv 0 \pmod{p}$ we are done.
If x is coprime to p then by Fermat's little Theorem we have

$$x^{ed} = x^{1+\phi k} = x^{1+(p-1)(q-1)k} = x \cdot (x^{(p-1)})^{(q-1)k} \equiv x \cdot 1 \pmod{p}$$

The same works for q, so $x^{ed} \equiv x \pmod{pq}$. □

Is RSA really safe? Can we get d from knowing n and e?

We know that $d \cdot e \equiv 1 \pmod{\phi}$ and ϕ we got from the prime factors of n: p and q. So if someone can find the factors of n then they can find d. However, factorising a number with two large (very large!) prime factors is extremely difficult.