Cryptography

Winter term 2004/05 by
Prof. Dr. Joachim Rosenthal,
University of Zürich

For personal use only

Felix Fontein

March 8, 2005

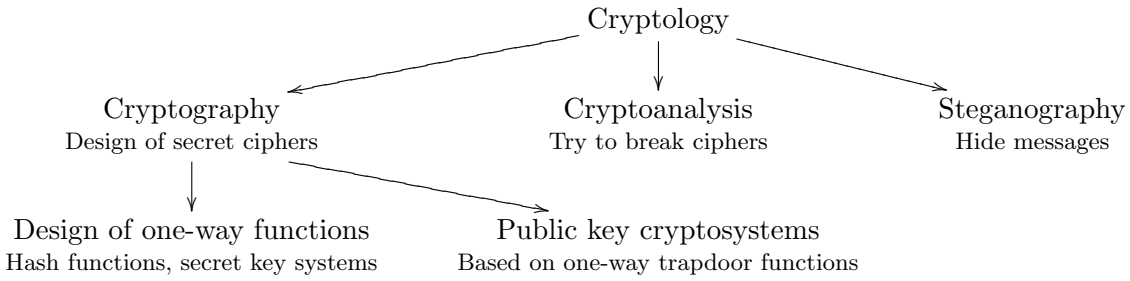
Contents

Chapter 1

Cryptography

1.1 Road Map to Cryptography

The area of cryptology contains lots of different subareas:



In this lecture, we will concentrate on cryptography. But what exactly is cryptography? We want to cite a definition from the *Handbook of Applied Cryptography* [MvOV96], the "bible" for applied cryptography:

Definition 1.1.1. Cryptography *is the study of mathematical techniques related to aspects of information security such as confidentiality in point-to-point communication, data integrity and authentification.*

Historical Remarks

- Around 1900 B.C., Egyptians used hyroglyphs to communicate secretly with their gods.
- The Romans used *Caesar ciphers*: By identifying the alphabet with $\mathbb{Z}_{26}$, that is the integers modulo 26, the cipher works by translating every letter by an offset, the secret key $k \in \mathbb{Z}_{26}$:

  $$\varphi : \mathbb{Z}_{26} \to \mathbb{Z}_{26}, \qquad m \mapsto m + k.$$

  This is a weak scheme, since by trying a maximum of 26 possibilities the plaintext can be found.
- Around 1600, *Vigenére* proposed the following improvement of the Caesar cipher: Instead of encrypting one letter at a time and using one key for all letters, his scheme encrypts n letters at a time, where each of them is translated by a (not necessary) different key:

  $$\varphi : \mathbb{Z}_{26}^n \to \mathbb{Z}_{26}^n, \qquad m \mapsto m + k \qquad \text{where } k \in \mathbb{Z}_{26}^n.$$

  This might look more complex than the Caesar cipher, but by employing statistical analysis like frequency analysis of letters, one can also defeat this scheme.
- In 1880, *Kerckhoff* formulated his principle:

  > "All the secrecy of a secret key system should rely on the secret key only."

- In 1917, Vernam proposed and received a patent for a Vigenére cipher where n goes to ∞, also called the *one time pad*. We will later see that the one time pad is provable secure. But it is not that useful in practice, since a key of at least the length of the message must be exchanged before. It is still used; it is rumoured that the Soviet and the U.S. governments exchanged lots of one time pad keys during the cold war, to be able to communicate absolutely secretly in emergency situations.
- In 1930, D. Hill proposed a system

  $$\varphi : \mathbb{Z}_{26}^n \to \mathbb{Z}_{26}^n, \qquad m \mapsto Am + k,$$

where $A \in GL_n(\mathbb{Z}_{26})$ and $k \in \mathbb{Z}_{26}^n$ form the key.[1] This is a weak scheme because of so called *plaintext attacks*: If the attacker knows a long enough sequence of pairs $(\tilde{m}_i, m_i)$ such that $\tilde{m}_i = Am_i + k$, he can compute A and k by employing basic linear algebra.

- In the Second World War, many new systems evolved. An example is the German *Enigma* machine.
- In 1949, C. Shannon published his article *Communication theory of secret systems*. He showed the existence of provable secure cyrptosystems.
- In 1976, Diffie and Hellmann realized the possibility of assymetric secret key systems, like
  - public key cryptography,
  - digital signatures and
  - zero knowledge proofs.

  Public key cryptosystems work as follows: If Alice wants to send a message to Bob, she looks up Bobs public key, which is publicy avaible. Then she encrypts the message with that key and sends it to Bob, who is the only person knowing the private key corresponding to the public key, and so can decrypt the message.

  The idea behind *digital signatures* is to mimic "real" signatures: Only one person can sign for a given identity, but everyone can check whether a signature belongs to that identity.

  An even more interesting concept are *zero knowledge proofs*: Alice wants to proof to Bob that she knows a secret, and at the end of the day Bob is convinced that Alice knows the secret, but has gained no clue about the secret itself.

[1] With $GL_n(R)$ we denote the invertible $n \times n$-matrices over a ring R. Also note that $A \in R^{n\times n}$ is invertible if and only if its determinant is a unit in R, i. e. $\det A \in R^*$.

This can be shown as follows: If A is invertible, then $1 = \det I_n = \det(AA^{-1}) = \det A \cdot \det A^{-1}$, so $\det A \in R^*$. Conversely, if $\det A \in R^*$, then since $AA^{\#} = \det A \cdot I_n$, we get $A^{-1} = (\det A)^{-1}A^{\#}$. (Here $A^{\#}$ is the adjoint matrix of A.)

Furthermore, note that an element $x \in \mathbb{Z}_n$ is invertible if and only if $\gcd(x, n) = 1$, i. e. if x and n are coprime. This can be proven by using the Bezout identity.

1.2 Introduction to Secret Key Systems

Definition 1.2.1. *Let X and Y be arbitrary sets. A function $\varphi : X \to Y$ is called a* one-way function *if $\varphi(x)$ can be effectively computed for every $x \in X$, and it is practically not possible to compute $x \in \varphi^{-1}(y)$ for almost all $y \in \operatorname{Im}\varphi$.*

Examples 1.2.2.

(1) *Let G be a finite group with $|G| \geq 2^{100}$ and $e \in \mathbb{N}$, for example $e = 17$. Also efficient multiplication should be possible. Define*

$$\varphi : G \to G, \qquad g \mapsto g^e.$$

Such functions are called of RSA type. *This is a good one-way function if $|G|$ is unknown! If $n = |G|$ is known, then by Lagrange we have $g^n = 1_G$ for all $g \in G$. If e and n are coprime, the* extended Euclidean algorithm *delivers a Bezout equation*

$$ed + nb = 1 \qquad \text{with } d, b \in \mathbb{Z}.$$

Then we have

$$\varphi(g)^d = (g^e)^d = g^{ed} = g^{1-nb} = g(g^n)^b = g1_G^b = g.$$

If n and e are not coprime, with the same method one can recover $g^{\gcd(n,e)}$ from g^e, but in general not g itself, since φ is not one-one, i. e. not injective.

(2) *Let $G = \langle g \rangle$ be a cyclic group with generator g, and $|G| \geq 2^{100}$. Assume again that multiplication in G is efficient. Let*

$$\varphi : \mathbb{Z} \to G, \qquad m \mapsto g^m.$$

As a notation: If $h = g^m$, we call m the discrete logarithm of h with base g, *written $m = \log_g h$. It is important to note that similar to the complex logarithm, the discrete logarithm is multi-valued, as for example $g^m = g^{m+|G|}$. For many groups, the* discrete log problem *(DLP) "given h and g, compute $\log_g h$" is considered a very hard problem.*

(3) *We want to define a one-way function $\varphi : X \to Y$, where $X = Y = \mathbb{Z}_2^{64}$. This scheme mimics the methods used by secret ciphers like Rijndael, the cipher behind AES. Consider the following multiplications on $\mathbb{Z}_2^{64}$:*

   (a) *The classic componentwise multiplication by interpreting $\mathbb{Z}_2^{64}$ as the 64-fold direct sum of $\mathbb{Z}_2$; we will denote this multiplication by $\otimes$.*

   (b) *By interpreting $\mathbb{Z}_2^{64}$ as $\mathbb{Z}_{2^{64}}$, for example by the bijection $(a_i)_i \mapsto \sum_i a_i 2^{i-1}$, one can define a $\mathbb{Z}_{2^{64}}$-like multiplication on $\mathbb{Z}_2^{64}$. We will denote this by $\cdot$.*

   (c) *Another way to interpret $\mathbb{Z}_2^{64}$ is by selecting a $\mathbb{F}_2$-basis of $\mathbb{F}_{2^{64}}$ and by this defining a mapping between the two spaces; we will denote the $\mathbb{F}_{2^{64}}$-multiplication on $\mathbb{Z}_2^{64}$ by $\times$.*

   (d) *Consider the mapping*

$$(x_i)_i \mapsto \begin{pmatrix} x_1 & \cdots & x_8 \\ x_9 & \cdots & x_{15} \\ \vdots & \ddots & \vdots \\ x_{57} & \cdots & x_{64} \end{pmatrix} \in \mathbb{Z}_2^{8\times 8}.$$

   We denote the $\mathbb{Z}_2^{8\times 8}$-multiplication on $\mathbb{Z}_2^{64}$ by $\circ$.

Given an $x \in X$, the cipher works by first doing a key expansion*:*

$$x_0 := x, \qquad x_{t+1} := x_t \cdot x_t + (x_t \circ x_t) \otimes x_t + x_t \times x_t \quad \text{for } t = 0, \ldots, 3.$$

Then, the one-way function φ can for example be defined like

$$\varphi(x) = x_1 \circ x_5 + (x_2 \otimes x_3) \cdot x_4.$$

The security of this scheme lies in the fact that, though the multiplications on $\mathbb{Z}_2^{64}$, $\mathbb{Z}_{2^{64}}$, $\mathbb{F}_{2^{64}}$ and $\mathbb{Z}_2^{8\times 8}$ alone can be described algebraically very well, the mixing of these operations makes it very hard or even impossible to employ algebraic methods to compute the preimage of an image element.

In the following, we will assume that every kind of information one wants to send and/or encrypt can be stored as a sequence of one's and zero's, i. e. as an element of $\mathbb{Z}_2^n$ for some n depending on the message. Of course, by employing bijective functions to other sets, also other sets than $\mathbb{Z}_2^n$ can be used to store information.

We want to give two more applications of one-way functions:

(1) Password storage: For example on UNIX, the Data Encryption Standard (DES) cipher is used to transform a user's password (given in ASCII letters, where an ASCII letter corresponds to an element of $\mathbb{Z}_{2^8}$) into a garbled looking string. For an attacker who got a copy of the password file, it is computationally hard to compute a preimage of the encrypted passwords.

(2) Hash functions: If X is a infinite set and Y finite, a one-way function $\varphi : X \to Y$ can be for example used to protect data against changes by computing the hash value for a big file; if the file is changed, for example by a malicious attacker or even by a hardware failure, recomputing the hash value will give with a high probability another value.

A more sophisticated version of the one-way functions are the *one-way functions with a secret key*: Let M, K and C be sets; we will call M the *message space*, K the *key space* and C the *cipher space*.

Definition 1.2.3. *A* secret key system *consists of maps*

$$\varphi : M \times K \to C \qquad \textit{and} \qquad \psi : C \times K \to M$$

called encryption *and* decryption*, such that*

(i) *for all $m \in M$ and all $k \in K$, we have $\psi(\varphi(m,k),k) = m$, and*

(ii) *for a fixed $m \in M$, the function $\varphi_m : k \mapsto \varphi(m,k)$ is a one-way function.*

Famous examples are

- the Enigma machine;
- the 1975 Data Encryption Standard (DES);
- the 2001 Advanced Encryption Standard (AES).

This still leaves open the question of how to exchange the secret key for communication, since when the attacker knows the key, everything is lost.

1.3 One-way Trapdoor Functions and the RSA System

In 1976, Diffie and Hellmann realized the importance of one-way trapdoor functions:

Definition 1.3.1. *A* one-way trapdoor function *is a one-way function $f : X \to Y$ having two additional properties:*

(i) *the function φ is one-one (injective), and*

(ii) *the designer has a trapdoor which allows him to efficiently compute $\varphi^{-1} : \operatorname{Im}\varphi \to X$.*

If one has such a function, it can be applied for example as follows:

(1) Secret key exchange: Alice publishes a one-way trapdoor function $\varphi : X \to Y$. Bob wants to send $k \in X$ to Alice, which should serve as the secret key for a symmetric encryption scheme. Instead of k he sends $\varphi(k)$ to Alice, which makes it impossible for an eavesdropper to get hold of k, but allows Alice to compute k by exploiting the trapdoor.

(2) Digital signatures: Alice deposits a one-way trapdoor function $\varphi : X \to Y$ with a trusted party; this could for example be a government institution. A signature would be for example

$$\varphi^{-1}(\text{“Alice, Zürich 21. October 2004”}).$$

It can be verified by applying φ to the signature; the idea behind this scheme is that no other person but Alice, the designer of the one-way trapdoor function, can compose a signature $x \in X$ such that $\varphi(x)$ gives a string as *“Alice, Zürich 21. October 2004”*.

This emphasizes that such a one-way trapdoor function could be very useful, but it does not helps coming up with such a function. In 1978, Rivest, Shamir and Adleman proposed the *RSA system*, which was the first instanciation of a one-way trapdoor function. The idea behind it is as follows: The designer (Alice) constructs a finite group G, where only Alice can compute $\phi := |G|$. As usual, it should be easy to do multiplication in G. In addition, Alice chooses an $e \in \mathbb{N}$ such that e and ϕ are coprime. Then

$$\varphi : G \to G, \qquad g \mapsto g^e$$

is a one-way trapdoor function.

Remarks 1.3.2.

(1) *The mapping φ is one-one. This follows directly from the next:*

(2) *The extended Euclidean algorithm delivers some $d \in \mathbb{Z}$ such that $ed + b\phi = 1$, where $b \in \mathbb{Z}$. Then we have*

$$\varphi^{-1} : G \to G, \qquad h \mapsto h^d,$$

since

$$(g^e)^d = g^{ed} = g^{1-b\phi} = g(g^\phi)^{-b} = g.$$

(3) *In RSA, one choses $G = \mathbb{Z}_n^*$, where n is the product of two large distinct primes p and q.*

Definition 1.3.3. *For a natural number $n \in \mathbb{N}_{>0}$, define the* Euler ϕ-function *as follows:*

$$\phi : \mathbb{N}_{>0} \to \mathbb{N}, \qquad n \mapsto \left|\mathbb{Z}_n^*\right|.$$

The next theorem will show how we can compute $\phi(n)$, if $n = pq$ and p, q are known.

Theorem 1.3.4. *If $n = \prod_{i=1}^k p_i^{e_i} \in \mathbb{N}_{>0}$, where the p_i are pairwise distinct primes and $e_i \in \mathbb{N}_{>0}$, then*

$$\phi(n) = \prod_{i=1}^k p_i^{e_i-1}(p_i - 1) = n \prod_{i=1}^k \frac{p_i - 1}{p_i}.$$

We will give two proofs of this theorem, one using elementary combinatorics and one employing the Chinese Remainder Theorem.

Proof #1. We will only show the case $e_i = 1$ here, i. e. $n = p_1 \cdots p_k$, by employing the inclusion/exclusion principle. Define

$$A_i := \{a \in \mathbb{Z}_n \mid p_i \text{ divides } a\}.$$

It is easy to see that

$$\mathbb{Z}_n^* = A_1^c \cap \cdots \cap A_k^c,$$

where $A_i^c = \mathbb{Z}_n \setminus A_i$. This gives

$$\begin{aligned}
\phi(n) = |\mathbb{Z}_n^*| &= |(A_1 \cup \cdots \cup A_k)^c| = n - |A_1 \cup \cdots \cup A_k| \\
&= n - \sum_i |A_i| + \sum_{i<j} |A_i \cap A_j| - \sum_{i<j<k} |A_i \cap A_j \cap A_k| \pm \cdots + (-1)^k \left| \bigcap_i A_i \right| \\
&= n - \sum_i \frac{n}{p_i} + \sum_{i<j} \frac{n}{p_i p_j} - \sum_{i<j<k} \frac{n}{p_i p_j p_k} \pm \cdots + (-1)^k \\
&= n(1 - \tfrac{1}{p_1}) \cdots (1 - \tfrac{1}{p_k}) = (p_1 - 1) \cdots (p_k - 1).
\end{aligned}$$

□

For the second proof, which works for all $n \in \mathbb{N}_{>0}$, we need the *Chinese Remainder Theorem* (CRT):

Theorem 1.3.5 (Chinese Remainder Theorem). *Let $n_1, \ldots, n_k \in \mathbb{N}_{>0}$ be pairwise coprime integers, and let $n = n_1 \cdots n_k$. Then*

$$\mathbb{Z}_n \cong \mathbb{Z}_{n_1} \times \ldots \mathbb{Z}_{n_k}.$$

Proof of the Chinese Remainder Theorem. Let $n = p_1^{e_1} \cdots p_k^{e_k}$ with p_i pairwise distinct primes and $e_i \in \mathbb{N}_{>0}$. We will show the case where $n_i := p_i^{e_i}$; the general case follows directly from this one.

Define the function

$$\varphi : \mathbb{Z} \to \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}, \qquad a \mapsto (a + n_1\mathbb{Z}, \ldots, a + n_k\mathbb{Z}).$$

It is clear that φ is a ring morphism, and one directly sees that

$$\ker \varphi = \bigcap_i \ker(x \mapsto x + n_i\mathbb{Z}) = \bigcap_i n_i\mathbb{Z} = n\mathbb{Z},$$

since n is the least common multiple of the n_i. By the isomorphism theorem, we have

$$\mathbb{Z}/n\mathbb{Z} = \mathbb{Z}/\ker\varphi \cong \operatorname{Im}\varphi \subseteq \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}.$$

We will show that φ is surjective, which completes the proof. Since $\mathbb{Z}/n\mathbb{Z} \cong \operatorname{Im}\varphi$, it is $|\operatorname{Im}\varphi| = |\mathbb{Z}/n\mathbb{Z}| = n$. Now we also have $|\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}| = \prod_{i=1}^k n_i = n$, and since n is finite, we get $\operatorname{Im}\varphi = \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$. □

For $n = n_1 \cdots n_k \in \mathbb{Z}$, where the n_i are relatively coprime, the Chinese Remainder Theorem gives

$$\mathbb{Z}_n \cong \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k},$$

which implies

$$\mathbb{Z}_n^* \cong \mathbb{Z}_{n_1}^* \times \cdots \times \mathbb{Z}_{n_k}^*.$$

Therefore, we get the following corollary of the Chinese Remainder Theorem:

Corollary 1.3.6. *If $n_1, \dots, n_k \in \mathbb{Z}$ are pairwise coprime and $n = n_1 \cdots n_k$, it is*

$$\phi(n) = \prod_{i=1}^{k} \phi(n_i).$$

Proof #2 of Theorem 1.3.4. By the above corollary, we get

$$\phi(n) = \prod_{i=1}^{k} \phi(p_i^{e_i}).$$

Now let us take a look at the case $n = p^e$ with p a prime and $e \in \mathbb{N}_{>0}$. Since $\gcd(a, p^e) = 1$ if and only if $\gcd(a, p) = 1$, we get

$$\left|\mathbb{Z}_{p^e}^*\right| = p^e - p^{e-1} = p^{e-1}(p-1).$$

□

Another very useful and easy to get corollary from the Chinese Remainder Theorem is the following:

Corollary 1.3.7 (Simultanous congruences). *Let $n = n_1 \cdots n_k$, where $n_1, \dots, n_k$ are pairwise coprime. Then for every $x_1, \dots, x_k \in \mathbb{Z}$ there exists a unique integer $\bar{x} \in \mathbb{Z}$ such that $0 \leq \bar{x} < n$ and*

$$\bar{x} \equiv x_i \pmod{n_i} \qquad \text{for every } i \in \{1, \dots, k\}.$$

Proof. The Chinese Remainder Theorem gives a unique $\bar{x} \in \mathbb{Z}_n$ such that

$$\bar{x} = \varphi^{-1}(x_1, \dots, x_k),$$

where φ is the function from the proof of the Chinese Remainder Theorem. □

Examples 1.3.8.

(1) *For the system $\bar{x} \equiv 1 \pmod 3$, $\bar{x} \equiv 3 \pmod 5$, one can easily see that $\bar{x} = 13$.*

(2) *Given the system $\bar{x} \equiv 13 \pmod{151}$, $\bar{x} \equiv 31 \pmod{131}$, it is not so obvious what the solution is. Euclids algorithm gives $a, b \in \mathbb{Z}$ such that $a \cdot 151 + b \cdot 131 = 1$. In this example, we get $a = 59$ and $b = 68$. Now $\bar{x} = 31 \cdot (59 \cdot 151) + 13 \cdot (68 \cdot 131) \mod 151 \cdot 131$; can you think of why this is the solution, and how to generalize this to more than two equations?*

Now, back to the RSA system. For *setting up* the system, Alice has to do the following:

(1) Alice choses two distinct primes $p, q \geq 10^{100}$.

(2) Alice computes $n = pq$ and $\phi(n) = (p-1)(q-1)$.

(3) Alice picks an $e \in \mathbb{N}$, $e < \phi(n)$, which is coprime to n, and computes $d \in \mathbb{N}$, $d < \phi(n)$ such that $ed + b\phi(n) = 1$ for some $b \in \mathbb{Z}$.

Now Alice *publishes* $\varphi : \mathbb{Z}_n \to \mathbb{Z}_n$, $m \mapsto m^e$. The information pieces p, q, $\phi(n)$ and d are kept secret by her.

Questions and remarks 1.3.9.

(1) *How difficult is it to find p and q from the public information? How difficult is it to factor a number $n \in \mathbb{N}$?*

(2) *Clearly if Bob sends $m \in \mathbb{Z}_n^*$, then Alice can decrypt it, i. e. compute m from m^e by exponentiating by d. But what happens if $m \in \mathbb{Z}_n \setminus \mathbb{Z}_n^*$?*

(It can be shown that decryption still works by using the Chinese Remainder Theorem. Can you figure out how to prove that?)

(3) *Knowing p and q is equivalent to knowing n and $\phi(n)$.*

(4) *How hard is it to compute m^e and c^d?*

To answer the fourth question, assume that m and e are random numbers in between $\{1, \ldots, n\}$. Use consecutive squaring to compute

$$m,\ m^2,\ m^4 = (m^2)^2,\ m^8 = (m^4)^2,\ m^{16} = (m^8)^2,\ \ldots,\ m^{2^k} = (m^{2^{k-1}})^2,$$

where[2] $k := \lfloor \log_2 n \rfloor$. Then m^e can be computed in at most $2k$ multipliations in $\mathbb{Z}_n$ as follows: Write e in binary representation, i. e.

$$e = \sum_{i=0}^{k} e_i 2^i, \qquad \text{where } e_i \in \{0, 1\}.$$

Then

$$m^e = \prod_{i=0}^{k} m^{e_i 2^i} = \prod_{\substack{i=0 \\ e_i \neq 0}}^{k} m^{2^i}.$$

Example 1.3.10. *Consider $e = 17$, that is $e = 2^0 + 2^4$. Thus we get*

$$m^{17} = m^{2^0 + 2^4} = m(((m^2)^2)^2)^2.$$

So computing m^e costs $\mathcal{O}(\log^3 n)$ bit operations, where $\mathcal{O}$ is described in the following short section:

[2]For a real number $x \in \mathbb{R}$, define $\lfloor x \rfloor := \max\{z \in \mathbb{Z} \mid z \leq x\}$ (floor) and $\lceil x \rceil := -\lfloor -x \rfloor$ (ceiling).

1.4 A Small Background in Complexity Theory

One writes $f(x) = \mathcal{O}(g(x))$ for $f, g : \mathbb{R} \to \mathbb{R}$ if there are constants $x_0, c \in \mathbb{R}$ such that $f(X) \leq cg(x)$ for all $x \geq x_0$. This is called the *big-$\mathcal{O}$ notation*. If $g \geq 0$, one has

$$\limsup_{x\to\infty} \frac{f(x)}{g(x)} < \infty \Longleftrightarrow f(x) = \mathcal{O}(g(x)).$$

Example 1.4.1. *The number of bit operations for adding two numbers $a, b \leq n$ is $\mathcal{O}(\log n)$, since the binary representation of a, b has at most length $\log n$. Similarly, multiplying two numbers $a, b \leq n$ requires $\mathcal{O}(\log^2 n)$ bit operations, if schoolbook multiplication is used. By employing more sophisticated methods, for example discrete Fourier transformations, multiplication can be made a lot faster for large n.*

Definition 1.4.2. *Given an algorithm for computing $f : \mathbb{N}^s \to \mathbb{R}$, $(a_1, \ldots, a_s) \mapsto f(a_1, \ldots, a_s)$, one says the algorithm has* polynomial time *if the number of bit operations is $\mathcal{O}(\log^k n)$ for some $k \in \mathbb{N}$, whenever $a_1, \ldots, a_s \leq n$. An algorithm which requires at least n^α bit operations for some $\alpha > 0$ is called an* exponential time *algorithm.*

In cryptography, problems for which polynomial time algorithms do exist are considered easy, while algorithms for which only exponential time algorithms do exist are considered (possibly) hard.

Definition 1.4.3. *A problem $\mathcal{P}$ is called a* polynomial time problem *once one knows a polynomial time algorithm for solving $\mathcal{P}$. All these problems form the class P.*

Examples 1.4.4.

(1) *Multiplying two numbers in $\mathbb{Z}_n$ is a polynomial time problem, thus it is in P.*

(2) *As was shown in [AKS02], the problem PRIMES (is a given number n prime?) is in P. More information about primality testing can be found in the next section.*

Definition 1.4.5. *A decision problem $\mathcal{P}$ is said to be in the class NP (*nondeterministic polynomially*), if*

(i) *the problem can be solved for someone with infinite computing power;*

(ii) *the answer can be verified in polynomial time.*

Example 1.4.6. *The problem FACTORING is clearly in NP, since once the factors are provided checking whether their product is the original number can be acomplished in polynomial time.*

Definition 1.4.7. *A decision problem $\mathcal{P}_1$* reduces *to a decision problem $\mathcal{P}_2$ if for any instance of $\mathcal{P}_1$ there is a polynomial time algorithm translating the problem to an instance of $\mathcal{P}_2$.*

Definition 1.4.8. *A decision problem $\mathcal{P}$ is called NP-*hard *if every other decision problem in NP reduces to $\mathcal{P}$. If moreover $\mathcal{P}$ is in the class of NP problems, one says that $\mathcal{P}$ is a NP-*complete *problem.*

Examples 1.4.9.

(1) *The traveling salesman problem.*

(2) *The subset sum problem (see later).*

(3) *The knapsack problem (see later).*

Remark 1.4.10. *A big open question in complexity theory is whether $P = NP$ or $P \subsetneqq NP$.*

1.5 Finding Primes and Primality Checking

For the RSA cryptosystem, one needs to construct two primes $p, q \geq 10^{100}$. How can this be done?

Remark 1.5.1. *There are infinitely many primes, as a simple argument shows: Assume $p_1, \ldots, p_n$ are all primes. Then, consider $p_1 \cdots p_n + 1$; none of the $p_1, \ldots, p_n$ divides this number, so it must contain another prime factor, a contradiction!*

A more interesting question is how the primes are distributed. This is partially answered by the following theorem:

Theorem 1.5.2 (Prime Number Theorem). *Let $\pi(x)$ denote the number of primes in the interval $[0, x]$. Then one has*

$$\lim_{x \to \infty} \frac{\pi(x)}{x/\log x} = 1.$$

This theorem has an important consequence: The chance that a randomly chosen integer with 100 digits is prime is roughly

$$\frac{10^{100}/\log 10^{100}}{10^{100}} = \frac{1}{100 \log 10} \approx \frac{1}{230}.$$

This leads to the following

Algorithm 1.5.3.

(1) *Pick a 100-digit number m not divisible by small primes like $2, 3, 5, \ldots$*

(2) *Check whether m is prime.*

(3) *If m is not prime, go back to step 1.*

This opens up another question: How to check whether a number $m \in \mathbb{N}$ is prime? One could try all possible divisors from 2 up to $\lfloor \sqrt{m} \rfloor$. The cost of that is $\mathcal{O}(m^{1/2} \log^2 m)$ bit operations: This is an exponential time algorithm!

In order to check if m is possibly prime, there are several probabilistic and deterministic algorithms which outperform this primitive algorithm a lot, i.ė. they are polynomial time. We will present three probabilistic algorithms and one deterministic one, which was published in the 2002 paper "PRIMES is in P" by three Indian computer scientists [AKS02]. The three probabilistic algorithms are:

(A) Fermat's test;

(B) Solovay-Strassen test;

(C) Miller-Rabin test.

1.5.1 The Fermat Test

We want to recite the *Little Fermat Theorem* for integers:

Theorem 1.5.4 (Little Fermat). *Let p be a prime and a an integer not divisible by p. Then*

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

Proof. It is $\left|\mathbb{Z}_p^*\right| = \phi(p) = p - 1$, and further we have $a \in \mathbb{Z}_p^*$; so by Lagrange this theorem follows. □

If p is not a prime, for some $a \in \mathbb{Z}_p^*$ this is often not the case. To be more precise about this, we first need a definition:

Definition 1.5.5. *For $n \in \mathbb{N}$, let*

$$U_n := \{a \in \mathbb{Z}_n^* \mid a^{n-1} \equiv 1 \pmod{n}\}.$$

Lemma 1.5.6. *For all $n \in \mathbb{N}$, the set U_n is a subgroup of $\mathbb{Z}_n^*$.*

Proof. Since $\mathbb{Z}_n^*$ is finite, it is enough to check that $ab \in U_n$ if $a, b \in U_n$. Indeed, if $a, b \in U_n$, then

$$(ab)^{n-1} = a^{n-1}b^{n-1} \equiv 1 \pmod{n}.$$

□

This implies that if $U_n \subsetneq \mathbb{Z}_n^*$, then by Lagrange we have $|U_n| \leq \frac{1}{2}|\mathbb{Z}_n^*| < \frac{1}{2}|\mathbb{Z}_n| = \frac{n}{2}$. Thus the probability that a randomly chosen $a \in \mathbb{Z}_n^*$ fulfills $a^{n-1} \equiv 1 \pmod{n}$ is at most $\frac{1}{2}$ in this case. This suggests the following algorithm, which is known as the *Fermat pseudoprime test*:

(1) Pick a canidate prime m.

(2) Check that m is not divisible by small primes.

(3) Pick random integers $a_1, \ldots, a_s \in \{1, \ldots, n-1\}$ and check whether $a_i^{n-1} \stackrel{?}{\equiv} 1 \pmod{n}$.

If $a_i^{n-1} \not\equiv 1 \pmod{n}$ for one i, then m is not prime by Little Fermat. If all tests succeed, then m is not neccessary prime! But if s is small, the probability that m is prime is larger than $1 - 2^{-s}$. Unfortunately, this probability cannot be send to one by increasing s up to infinity, for the following reasons:

Definition 1.5.7. *A number n which is not prime is called a* Carmichael number *if $U_n = \mathbb{Z}_n^*$, that is for all $a \in \mathbb{Z}_n^*$ we have $a^{n-1} \equiv 1 \pmod{n}$.*

Carmichael numbers do exist, the smallest one is 561. Before characterizing them further, we would like to point out that there even exist infinitely many of them.

Theorem 1.5.8. *Let $n \in \mathbb{N}$.*

(a) *If p is a prime and p^2 divides n, then n is not Carmichael. Thus all Carmichael numbers are squarefree.*

(b) *If n is composite, odd and squarefree, then n is Carmichael if and only if $p \mid n$ implies $(p-1) \mid (n-1)$.*

(c) *If n is Carmichael, then n has at least three prime factors.*

Proof.

(a) Write $n = p^e m$ where $\gcd(p, m) = 1$, and assume $e \geq 2$. By the Chinese Remainder Theorem, we have

$$\mathbb{Z}_n^* \cong \mathbb{Z}_{p^e}^* \times \mathbb{Z}_m^*.$$

The order of $\mathbb{Z}_{p^e}^*$ is $p^{e-1}(p-1)$, so p divides $\phi(p^e)$. By Sylow, there is an element $a \in \mathbb{Z}_{p^e}^*$ of order p. So there is some $b \in \mathbb{Z}_n^*$ which corresponds to $(a, 1) \in \mathbb{Z}_{p^e} \times \mathbb{Z}_m$; and b also has order p.

Now, it must be $b^{n-1} \not\equiv 1 \pmod{n}$, since elsewise p divides $n-1$, but since p already divides n this is a contradiction.

(b) Assume $n = p_1 \cdots p_s$, where the p_i are distinct odd primes. By the Chinese Remainder Theorem,

$$\mathbb{Z}_n^* \cong \mathbb{Z}_{p_1}^* \times \cdots \times \mathbb{Z}_{p_s}^*.$$

Chose some $x \in \mathbb{Z}_n^*$, and let x correspond to $(x_1 \ldots, x_s)$. Then $x^{n-1} \equiv 1 \pmod{n}$ if and only if $x_i^{n-1} \equiv 1 \pmod{p_i}$ for $i = 1, \ldots, s$. So if $(p_i - 1) \mid (n-1)$ for all i, this is always the case, which completes the 'if' part of the proof.

For the 'only if' part, assume there is an i such that $(p_i - 1) \nmid (n-1)$. Let $a \in \mathbb{Z}_{p_i}^*$ be a primitive element, that is a generates $\mathbb{Z}_{p_i}^*$. Then $a^{n-1} \not\equiv 1 \pmod{p_i}$. So if $b \in \mathbb{Z}_n^*$ corresponds to $(1, \ldots, 1, a, 1, \ldots, 1)$, then $b^{n-1} \not\equiv 1 \pmod{n}$ by the Chinese Remainder Theorem. Thus n cannot be Carmichael.

(c) Assume $n = pq$, where p and q are primes and $p > q$. If n would be Carmichael, by (b) we get $(p-1) \mid (n-1)$, and hence there is an $\lambda \in \mathbb{N}$ such that $\lambda(p-1) = n-1 = pq-1$. This means

$$q - \lambda = \frac{\lambda p - \lambda + 1}{p} - \lambda = \frac{1-\lambda}{p} \in \mathbb{N},$$

which implies $p \mid (\lambda - 1)$ and so $\lambda \geq p+1$. Thus,

$$n - 1 = \lambda(p-1) \geq (p+1)(p-1) = p^2 - 1 > pq - 1 = n - 1,$$

a contradiction.

□

As a result, the following can be said: If $n \in \mathbb{N}$ is a number, there are two possibilities:

- $U_n = \mathbb{Z}_n^*$, which happens if and only if n is prime or Carmichael;
- $U_n \subsetneq \mathbb{Z}_n^*$, which happens if and only if n is composite and $[\mathbb{Z}_n^* : U_n] \geq 2$.

Thus for a number $n \in \mathbb{N}$ which is neither prime nor Carmichael, the chance that a random $a \in \mathbb{Z}_n$ fails $a^{n-1} \equiv 1 \pmod{n}$ (and thus proves that n is not prime) is at least $\frac{1}{2}$.

1.5.2 The Solovay-Strassen Test (1977)

Before we can present the results by Solovay and Strassen, we first have to introduce some results from elementary number theory.

Definition 1.5.9. *Let $\mathbb{F}$ be a finite field. An element $u \in \mathbb{F}^* = \mathbb{F} \setminus \{0\}$ is called a* quadratic residue *if the equation $x^2 = u$ has a solution in $\mathbb{F}$. Otherwise, u is called a* quadratic nonresidue.

Example 1.5.10. *Let $\mathbb{F} = \mathbb{Z}_{11}$ and take a look at the following table:*

x	1	2	3	4	5	6	7	8	9	10
x^2	1	4	9	5	3	3	5	9	4	1

So $\{1, 3, 4, 5, 9\}$ are the quadratic residues of $\mathbb{Z}_n$.

In this example one can already get an idea what happens in a finite field: Both $-x$ and x are mapped onto the same number x^2 by squaring, and thus (if $x \neq -x$ for all $x \in \mathbb{F}^*$) at most half of the elements can be quadratic residues. The following lemma gives a more exact result:

Lemma 1.5.11. *When the characteristic* $\operatorname{Char}\mathbb{F} = 2$*, then every element of $\mathbb{F}^*$ is a quadratic residue. If* $\operatorname{Char}\mathbb{F} \neq 2$ *then exactly half the elements of $\mathbb{F}^*$ are quadratic residues.*

Proof. Consider the squaring map $SQ : \mathbb{F} \to \mathbb{F}$, $x \mapsto x^2$. If $\operatorname{Char}\mathbb{F} = 2$, then SQ is a $\mathbb{Z}_2$-linear map. Further $\ker SQ = \{0\}$, and thus SQ is one-one. Since $\mathbb{F}$ is finite, SQ must also be onto (surjective). Since $SQ(\mathbb{F}^*)$ are the quadratic residues of $\mathbb{F}$ we are done.

If $\operatorname{Char}\mathbb{F} \neq 2$, then $SQ(a) = SQ(b)$ if and only if $a = -b$ or $a = b$. Since the only $x \in \mathbb{F}$ satisfying $x = -x$ is $x = 0$, every quadratic residue corresponds exactly to two elements of $\mathbb{F}^*$. This completes the proof. □

At first, we want to consider $\mathbb{F} = \mathbb{Z}_p$ for a prime p.

Definition 1.5.12. *Let p be an odd prime and $a \in \mathbb{N}$ arbitrary. Then let*

$$\left(\frac{a}{p}\right) := \begin{cases} 0 & \text{if } a \equiv 0 \pmod{p}, \\ 1 & \text{if } a \text{ is a quadratic residue in } \mathbb{Z}_p, \\ -1 & \text{elsewise} \end{cases}$$

be the Legendre symbol.

Example 1.5.13. *It is* $\left(\frac{7}{11}\right) = -1$.

Theorem 1.5.14 (Euler, 1760). *If p is an odd prime and $a \in \mathbb{N}$, then*

$$a^{\frac{p-1}{2}} \equiv \left(\frac{a}{p}\right) \pmod{p}.$$

Proof. Assume $a \not\equiv 0 \pmod{p}$. Then $a^{\frac{p-1}{2}} \equiv \pm 1 \pmod{p}$, since the polynomial $x^2 = 1$ has exactly the two solutions ± 1 in $\mathbb{Z}_p$ and $a^{p-1} \equiv 1 \pmod{p}$ by Little Fermat. If a is a quadratic residue, there exists some $v \in \mathbb{Z}_p$ such that $v^2 \equiv a \pmod{p}$, and thus $a^{\frac{p-1}{2}} \equiv v^{p-1} \equiv 1 \pmod{p}$. Now consider the set

$$Q := \{x \in \mathbb{Z}_p^* \mid x^{\frac{p-1}{2}} \equiv 1 \pmod{p}\}.$$

By the Fundamental Theorem of Algebra, the polynomial $x^{\frac{p-1}{2}} - 1$ has at most $\frac{p-1}{2}$ roots, and thus $|Q| \leq \frac{p-1}{2}$. But we just have shown that Q contains at least $\frac{p-1}{2}$ elements, and thus Q must be exactly the set of quadratic residues. □

But in our case, we want to check whether or not n is prime. For that, we require a definition of what $\left(\frac{a}{n}\right)$ means if n is not prime (in this case $\mathbb{Z}_n$ is not a field), and how quadratic residues behave in $\mathbb{Z}_n$.

Definition 1.5.15. *Let $n \geq 0$ be an odd integer and $n = p_1^{e_1} \cdots p_s^{e_s}$, where the p_i are distinct primes. Then for $a \in \mathbb{N}$ let*

$$\left(\frac{a}{n}\right) := \left(\frac{a}{p_1}\right)^{e_1} \cdots \left(\frac{a}{p_s}\right)^{e_s}$$

be the Jacobi symbol.

Theorem 1.5.16. *Let $n \in \mathbb{N}$ and $a_1, a_2, a \in \mathbb{N}$.*

(1) *If $a_1 \equiv a_2 \pmod{n}$, then $\left(\frac{a_1}{n}\right) = \left(\frac{a_2}{n}\right)$.*

(2) *It is $\left(\frac{a_1 a_2}{n}\right) = \left(\frac{a_1}{n}\right)\left(\frac{a_2}{n}\right)$.*

(3) *The following* inversion formula*, which is also known as the* quadratic reciprocity law*, holds:*

$$\left(\frac{a}{n}\right) = \begin{cases} -\left(\frac{n}{a}\right) & \text{if } a \equiv n \equiv 3 \pmod{4}, \\ \left(\frac{n}{a}\right) & \text{otherwise.} \end{cases}$$

(4) *If n is odd, then*

$$\left(\frac{2}{n}\right) = \begin{cases} 1 & \text{if } n \equiv \pm 1 \pmod{8}, \\ -1 & \text{if } n \equiv \pm 3 \pmod{8}. \end{cases}$$

A remark for the proof: The statements (1) and (2) directly follow from the definition. Proofs for the other statements can be found for example in every book about elementary number theory which covers quadratic residues.

Remark 1.5.17. *The first statement can be interpretet such that the Jacobi symbol becomes a map from $\mathbb{Z}_n$ to $\{-1, 0, 1\}$; and the second says that $\left(\frac{\cdot}{n}\right) : \mathbb{Z}_n^* \to \{-1, 1\}$, $a \mapsto \left(\frac{a}{n}\right)$ is a group homomorphism.*

The theorem allows efficient computation of $\left(\frac{a}{n}\right)$ for large a, n:

Example 1.5.18. *It is*

$$\left(\frac{176}{221}\right) \underset{(2)}{=} \left(\frac{2}{221}\right)^4 \left(\frac{11}{221}\right) = \left(\frac{11}{221}\right) \underset{(3)}{=} \left(\frac{221}{11}\right) \underset{(1)}{=} \left(\frac{1}{11}\right) = 1.$$

Remark 1.5.19. *An algorithm can be deduced whose complexity is at most $\mathcal{O}(\log^3 n)$ bit operations.*

Theorem 1.5.20 (Solovay-Strassen). *Assume n is odd.*

(a) *The set*

$$V := \left\{ x \in \mathbb{Z}_n^* \mid x^{\frac{n-1}{2}} \equiv \left(\frac{x}{n}\right) \pmod{n} \right\}$$

is a subgroup of $\mathbb{Z}_n^*$.

(b) *It is* $V = \mathbb{Z}_n^*$ *if and only if* n *is prime.*

The consequence is that if n is not prime, then for at most half of the numbers $a \in \mathbb{Z}_n^*$ we have $a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n}$. Thus by randomly chosing t integers $a_1, \ldots, a_t \in \mathbb{Z}_n^*$, one checks whether a number n is not prime or prime with a probability at least $1 - \frac{1}{2^t}$. This test is called the *Solovay-Strassen test*.

The cost of the test (for a fixed t) is $\mathcal{O}(\log^3 n)$.

Proof of the Solovay-Strassen theorem.

(a) Again, it suffices to show $ab \in V$ if $a, b \in V$. So let $a, b \in V$, then we have

$$\left(\frac{a}{n}\right)\left(\frac{b}{n}\right) = \left(\frac{ab}{n}\right) \qquad \text{and} \qquad a^{\frac{n-1}{2}} b^{\frac{n-1}{2}} \equiv (ab)^{\frac{n-1}{2}} \pmod{n}.$$

(b) If n is prime, by Euler $V = \mathbb{Z}_n^*$. Otherwise, if n is not prime, let us assume $V = \mathbb{Z}_n^*$. Then $x^{n-1} \equiv 1 \pmod{n}$ for all $x \in \mathbb{Z}_n^*$; thus n has to be Carmichael, and $n = p_1 \cdots p_s$ where the p_i are pairwise distinct primes, and $s \geq 3$, and futhermore $p_i - 1$ divides $n - 1$ for every i by theorem 1.5.8. Consider the Chinese Remainder Theorem:

$$\mathbb{Z}_n^* \equiv \mathbb{Z}_{p_1}^* \times \cdots \times \mathbb{Z}_{p_s}^*.$$

Let $b \in \mathbb{Z}_{p_1}^*$ a quadratic nonresidue, and let $a \in \mathbb{Z}_n^*$ correspond to $(b, 1, \ldots, 1)$. Then $a^{\frac{n-1}{2}}$ corresponds to $(b^{\frac{n-1}{2}}, 1, \ldots, 1)$, and since the correspondence is one-to-one and $a^{\frac{n-1}{2}} \equiv \pm 1 \pmod{n}$ (because of $\mathbb{Z}_n^* = V$), it must be $a^{\frac{n-1}{2}} \equiv 1 \pmod{n}$.

On the other hand we have

$$\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \cdots \left(\frac{a}{p_s}\right) = \left(\frac{b}{p_1}\right)\left(\frac{1}{p_2}\right) \cdots \left(\frac{1}{p_s}\right) = \left(\frac{b}{p_1}\right) = -1,$$

contradicting $V = \mathbb{Z}_n^*$.

□

1.5.3 The Miller-Rabin Test

Now we want to present another probabilistic primality test, which is more efficient than the first two in the sense that the probability for a failure is at most $\frac{1}{4}$ for one round in the test, and not $\frac{1}{2}$. It is currently one of the most used tests for primality. But before we present that test, we again need some preparations.

Lemma 1.5.21. *Let* n *be prime and* $n - 1 = 2^s d$ *where* d *is odd. If* $a \in \mathbb{Z}_n^*$*, then either* $a^d \equiv 1 \pmod{n}$*, or there exists some* $r \in \{0, 1, \ldots, s-1\}$ *such that* $a^{2^r d} \equiv -1 \pmod{n}$.

Proof. Clearly $\operatorname{ord}(a)$ divides $n - 1 = |\mathbb{Z}_n^*|$. So $\operatorname{ord}(a^d) = 2^\ell$ for some $0 \leq \ell \leq s$. If $\ell = 0$, then $a^d \equiv 1 \pmod{n}$. Otherwise $(a^d)^{2^{\ell-1}} \not\equiv 1 \pmod{n}$ and $(a^d)^{2^\ell} \equiv 1 \pmod{n}$, and since 1 has only the two square roots ± 1 modulo n since n is odd, it must be $(a^d)^{2^{\ell-1}} \equiv -1 \pmod{n}$. □

Definition 1.5.22. *For some odd* $n \in \mathbb{N}$*, define the following sets:*

- *The* Fermat liars
$$F(n) := \{a \in \mathbb{Z}_n^* \mid a^{n-1} \equiv 1 \pmod{n}\};$$
- *The* Euler liars
$$E(n) := \left\{ a \in \mathbb{Z}_n^* \mid a^{\frac{n-1}{2}} \equiv \left(\frac{a}{n}\right) \pmod{n} \right\};$$

- *The* strong liars

$$S(n) := \{a \in \mathbb{Z}_n^* \mid a^d \equiv 1 \pmod{n} \text{ or } a^{2^r d} \equiv -1 \pmod{n} \text{ for some } r \in \{0, 1, \ldots, s-1\}\},$$

where $n - 1 = 2^s d$ *such that* d *is odd.*

Example 1.5.23. *Let* $n = 65$. *Then* $|\mathbb{Z}_n^*| = \phi(65) = 4 \cdot 12 = 48$.

- *It is*

$$F(65) = \{1, 8, 12, 14, 18, 21, 27, 31, 34, 38, 44, 47, 51, 53, 57, 64\}$$

a subgroup of index 3;

- *It is*

$$E(65) = \{1, 8, 14, 18, 47, 51, 57, 64\}$$

a subgroup of index 6;

- *It is*

$$S(65) = \{1, 8, 18, 47, 57, 64\};$$

this is not a subgroup, since $8 \cdot 18 \equiv 14 \pmod{65}$.

Theorem 1.5.24. *For all odd* n, *one has that*

$$S(n) \subseteq E(n) \subseteq F(n) \subseteq \mathbb{Z}_n^*.$$

Thus, $S(n) = \mathbb{Z}_n^*$ *if and only if* n *is prime.*

Proof. If n is prime, we have $S(n) = \mathbb{Z}_n^*$ by the lemma. So let n be composite. By Solovay-Strassen, $E(n) \subsetneq \mathbb{Z}_n^*$. Futhermore, it is clear that $E(n) \subseteq F(n) \subseteq \mathbb{Z}_n^*$. So we can complete the proof by showing $S(n) \subseteq E(n)$.

Assume $a \in S(n)$ and $n - 1 = 2^s d$, where d is odd. Let k be the smallest integer such that $a^{2^k d} \equiv 1 \pmod{n}$; by assumption we have $k \in \{0, 1, \ldots, s\}$. Assume $n = p_1^{e_1} \cdots p_t^{e_t}$, where the p_i are distinct primes.

We first take a look at the case $k = 0$. For every i we have $a^d \equiv 1 \pmod{p_i}$, and thus $\operatorname{ord}_{p_i} a$ divides d. Since d is odd, $\operatorname{ord}_{p_i} a$ must be odd. Further ord_{p_i} divides $p_i - 1$ and thus $a^{\frac{p_i-1}{2}} \equiv 1 \pmod{p_i}$, which implies $\left(\frac{a}{p_i}\right) = 1$ by Euler. But this means $\left(\frac{a}{n}\right) = 1 \equiv a^{\frac{n-1}{2}} \pmod{n}$, so we have $a \in E(n)$.

The second case is $k > 0$; in that case $a^{2^{k-1}d} \equiv -1 \pmod{n}$. For any i we have $a^{2^k d} \equiv 1 \pmod{p_i}$ and $a^{2^{k-1}d} \equiv -1 \pmod{p_i}$, and thus $\operatorname{ord}_{p_i} a$ divides $2^k d$, but not divides $2^{k-1} d$. So we can write $\operatorname{ord}_{p_i} a = 2^k d_i$, where d_i is odd. Since $\operatorname{ord}_{p_i} a$ divides $p_i - 1$, we know that 2^k divides $p_i - 1$. Thus we can write $p_i = 2^k b_i + 1$ where $b_i \in \mathbb{Z}$. Note that

$$a^{\frac{\operatorname{ord}_{p_i} a}{2}} \equiv -1 \pmod{p_i}.$$

Thus by Euler

$$\left(\frac{a}{p_i}\right) \equiv a^{\frac{p_i-1}{2}} \equiv a^{\frac{\operatorname{ord}_{p_i} a}{2} \cdot \frac{p_i-1}{\operatorname{ord}_{p_i} a}} \equiv (-1)^{\frac{p_i-1}{\operatorname{ord}_{p_i} a}}$$
$$\equiv (-1)^{\frac{p_i-1}{2^k d_i}} \equiv (-1)^{\frac{p_i-1}{2^k}} = (-1)^{b_i} \pmod{p_i},$$

since d_i is odd. Further we have

$$n = \prod_{i=1}^{t} p_i^{e_i} = \prod_{i=1}^{t} (2^k b_i + 1)^{e_i} \equiv \prod_{i=1}^{t} (1 + 2^k b_i e_i) \equiv 1 + 2^k \sum_{i=1}^{t} b_i e_i \pmod{2^{2k}}.$$

Therefore we have

$$2^{s-1} d = \frac{n-1}{2} \equiv 2^{k-1} \sum_{i=1}^{t} b_i e_i \pmod{2^k},$$

and thus

$$2^{s-k}d \equiv \sum_{i=1}^{t} b_i e_i \pmod 2.$$

So we finally get

$$\begin{aligned} a^{\frac{n-1}{2}} &= a^{2^{s-1}d} = (a^{2^{k-1}d})^{2^{s-k}} \equiv (-1)^{2^{s-k}} \equiv (-1)^{\sum_{i=1}^{t} b_i e_i} \\ &\equiv \prod_{i=1}^{t}((-1)^{b_i})^{e_i} \equiv \prod_{i=1}^{t}\left(\frac{a}{p_i}\right)^{e_i} = \left(\frac{a}{n}\right) \pmod n, \end{aligned}$$

and thus $a \in E(n)$. □

Theorem 1.5.25 (Miller and Rabin). *If n is odd and composite, then $|S(n)| \leq \frac{1}{4}\phi(n)$ except if $n = 9$; in that case $|S(n)| = 2$, while $\phi(n) = 6$.*

Proof. We distinguish two cases:

1. The first case is that n is Carmichael.

   Let $n = p_1 \cdots p_t$, where the p_i are distinct primes, and $p_i - 1$ divides $n-1$ for all i, and $t \geq 3$. (This can be assumed by theorem 1.5.8.) Define numbers $s_1, \ldots, s_t$ such that $n - 1 = 2^{s_i}(p_i - 1)d_i$, where d_i is odd for every i. Without loss of generality, we can assume $s_1 \leq \cdots \leq s_t$. Let $s := s_1 = \min\{s_1, \ldots, s_t\}$.

   Then $a^{\frac{n-1}{2^s}} \equiv 1 \pmod n$ for all $a \in \mathbb{Z}_n^*$, which one can easily see by applying the Chinese Remainder Theorem. Furthermore, $\frac{n-1}{2^s}$ is even.

   We distinguish two more cases:

   1a. The first is that $s = s_i$ for all i. Then $\frac{n-1}{2^{s+1}}$ is an odd multiple of $\frac{p_i-1}{2}$. Then $S(n)$ is contained in the subgroup

   $$A_1 := \{a \in \mathbb{Z}_n^* \mid a^{\frac{n-1}{2^{s+1}}} \equiv \pm 1 \pmod n\}.$$

   Let $a(k_1, \ldots, k_t)$ be the element in $\mathbb{Z}_n^*$ defined via

   $$\psi : \mathbb{Z}_n \to \mathbb{Z}_{p_1} \times \cdots \times \mathbb{Z}_{p_t}, \qquad a(k_1, \ldots, k_t) \mapsto (g_1^{k_1}, \ldots, g_t^{k_t}),$$

   where the g_i's are generators of the $\mathbb{Z}_{p_i}^*$'s. Then $a(k_1, \ldots, k_t)^{\frac{n-1}{2^{s+1}}} \equiv \pm 1 \pmod n$ if and only if either all k_i are even, or all k_i are odd. Since $t \geq 3$, then it follows that $|S(n)| \leq \frac{1}{2^{t-1}}\phi(n) \leq \frac{1}{4}\phi(n)$.

   1b. The second is $s_t > s$. Then $\frac{n-1}{2^{s+1}}$ is a multiple of $p_t - 1$, and hence even. So

   $$S(n) \subseteq A_0 := \{a \in \mathbb{Z}_n^* \mid a^{\frac{n-1}{2^{s+1}}} \equiv 1 \pmod n\}.$$

   Since it is $A_0 \neq \mathbb{Z}_n^*$, we know that $|A_0| \leq \frac{1}{2}\phi(n)$. Additionally, we have

   $$S(n) \subseteq A_2 := \{a \in \mathbb{Z}_n^* \mid a^{\frac{n-1}{2^{s+2}}} \equiv \pm 1 \pmod n\},$$

   which is clearly a subgroup of A_0.
   We now claim $A_2 \subsetneq A_0$; which again is left to be proved by the reader.
   Together it follows that $|S(n)| \leq |A_2| \leq \frac{1}{2}|A_0| \leq \frac{1}{4}\phi(n)$.

2. The second case is that n is not Carmichael.

   We know that $S(n) \subseteq F(n) \subsetneq \mathbb{Z}_n^*$ and $|F(n)| \leq \frac{1}{2}\phi(n)$.

   As an excercise, construct a subgroup $W \subseteq F(n)$ such that

   (i) $S(n) \subseteq W$ and
   (ii) $W \subsetneq F(n)$.

   Hint: Let $W = \{a \in \mathbb{Z}_n^* \mid a^{2^\ell d} \equiv \pm 1 \pmod n\}$ for some ℓ.

□

So let us sum this up: Let n be a canidate prime; for example, $n \approx 10^{100}$, then approximately $\frac{1}{230}$ of the numbers are prime. The probability is higher when small factors do not produce division.

Take random numbers $a_1, \ldots, a_t \in \mathbb{Z}_n^*$, and compute $n - 1 = 2^s d$ where d is odd. Then compute for $i = 1, \ldots, t$

$$a_i^d \stackrel{?}{\equiv} 1 \pmod{n} \qquad \text{and} \qquad a_i^{2^\ell d} \stackrel{?}{\equiv} -1 \pmod{n}, \quad \text{where } \ell = 0, \ldots, s-1.$$

If neither happens for a particular i, then we have proven that n is not prime by the first lemma of this subsection! If one of the cases happens for every i, then the likelihood that n is prime is at least $1 - 4^{-t}$ by Miller-Rabin.

In practice, take for example $t = 20$. This results in prime numbers with probability at least $1 - 2^{40} \approx 1 - 10^{-12}$. But what is the cost of this test? It is $\mathcal{O}(\log^3 n)$ bit operations, since we need $\mathcal{O}(\log n)$ multiplications in $\mathbb{Z}_n^*$.

This test is called the *Miller-Rabin pseudoprime test*, and it is probably the most-used non-deterministic test today: It does not bears the problems which the Fermat test has, and it includes the Solovay-Strassen test while being easier to compute, since $\left(\frac{a}{n}\right)$ does not needs to be evaluated.

1.5.4 Deterministic Primality Tests

Let us leave the area of non-deterministic tests and return to the deterministic ones. As we have seen, simply trying to divide by all possible prime factors is not a good idea, since it is an exponential time algorithm. For a long time, it was not clear if there exist deterministic polynomial time primality test. This question was answered positively in August 2002 by Agrawal, Kayal and Saxena, when they published a preprint of their paper [AKS02], which gives a polynomial time algorithm! Unfortunately, the complexity for the algorithm is quite high even though it is polynomial: The current version has a complexity of $\mathcal{O}(\log^{10.5} n)$, where the original version even had $\mathcal{O}(\log^{12} n)$. Thus, for practical applications where a primality test is required to be fast, non-deterministic algorithms are still in use.

In this subsection, we want to sketch the *idea of this paper*. Consider the polynomial ring $\mathbb{Z}_n[x]$.

Lemma 1.5.26. *For all $a \in \mathbb{Z}_n^*$, it is $(x + a)^n \equiv x^n + a \pmod{n}$ if and only if n is prime.*

Proof. If n is prime, one has $(x + y)^n = x^n + y^n$ in $\mathbb{Z}_n[x, y]$, and by Little Fermat, $a^n \equiv 1 \pmod{n}$ if $a \in \mathbb{Z}_n^*$.

If n is not prime, then a has to be Carmichael. It follows that $n = p_1 \cdots p_t$ and many binomial coefficients $\binom{n}{m}$ are non-zero modulo n. □

Remark 1.5.27. *If $n = pq$, where $p < q$ are primes, then*

$$(x + a)^n = x^n + 0 + \cdots + 0 + \binom{n}{p} x^{n-p} a^p + 0 + \ldots,$$

and $\binom{n}{p}$ is divisible by q.

The lemma cannot be used directly for practical reasons, since representing $(x+a)^n \mod n$ or even just computing it would be an exponential time algorithm! The idea of AKS is now to compute $(x + a)^n \pmod{n, x^r - 1}$ for several small r.

1.6 Finite Fields

In this section we want to recall several facts about finite fields which we will need later.

Proposition 1.6.1. *Let $\mathbb{F}$ be a finite field and $q = |\mathbb{F}|$. Then $q = p^n$ where $n \in \mathbb{N}_{>0}$ and p is prime. Further $\mathbb{Z}_p$ is contained in $\mathbb{F}$ as a subfield.*

Proof. Define a map $\psi : \mathbb{Z} \to \mathbb{F}$ as follows: Map $0 \mapsto 0$, $n \mapsto 1 + \cdots + 1$ (n times) and $-n \mapsto -(1+\cdots+1)$ (n times), where $n \in \mathbb{N}_{>0}$. It is easy to see that this is a ring homomorphism. Since $\mathbb{Z}$ is a principal ideal domain, $\ker \psi = m\mathbb{Z}$ for some $m \in \mathbb{N}$; and thus $\mathbb{Z}_m$ is embedded as $\psi(\mathbb{Z})$ in $\mathbb{F}$. Since $\mathbb{F}$ contains no zero divisors, $m\mathbb{Z}$ must be a prime ideal. In addition $\mathbb{F}$ is finite, and thus $m > 0$. So m must be prime. Now $\mathbb{F}$ is a $\mathbb{Z}_m$-vector space, and as $|\mathbb{F}| < \infty$ we have $n := \dim_{\mathbb{Z}_m} \mathbb{F} < \infty$, and thus $|\mathbb{F}| = |\mathbb{Z}_m^n| = m^n$, and we conclude since m is prime. □

Remark 1.6.2. *If $\mathbb{F}$ is an arbitrary (not neccessary finite) field, then the map ψ gives us the* characteristic *of $\mathbb{F}$:*

$$\operatorname{Char} \mathbb{F} = \begin{cases} 0 & \text{if } \ker \psi = 0, \\ p & \text{if } \ker \psi = p\mathbb{Z}. \end{cases}$$

Examples 1.6.3.

(a) *For $\mathbb{Q}$, $\mathbb{R}$ and $\mathbb{C}$, the characteristic is zero since they contain $\mathbb{Z}$ as a subring, and thus ψ is injective.*

(b) *Let $\mathbb{F} = \mathbb{Z}_2[x]/(x^3+x+1)$; from the exercises we know this is a field. We have $|\mathbb{F}| = 8 = 2^3$, and further $\mathbb{Z}_2 \subseteq \mathbb{F}$ and $\operatorname{Char} \mathbb{F} = 2$.*

Theorem 1.6.4. *For each prime p and $n \in \mathbb{N}_{>0}$ there exists a unique (up to isomorphism) field $\mathbb{F}$ such that $|\mathbb{F}| = p^n$.*

Proof. Consider $f = x^{p^n} - x \in \mathbb{Z}_p[x]$. Now $f' = -1$, and thus f has only simple roots. Let $\mathbb{K} \supseteq \mathbb{Z}_p$ be an extension field such that $f = \prod_{i=1}^{p^n}(x - x_i)$, where $x_i \in \mathbb{K}$ (for example, take the algebraic closure of $\mathbb{Z}_p$, or a splitting field of f). Let $\mathbb{F} := \{x_1, \ldots, x_{p^n}\}$. We will show that $\mathbb{F}$ is a field:

It is easy to see that $0, 1 \in \mathbb{F}$. If $x, y \in \mathbb{F}$, then $(x-y)^{p^n} = x^{p^n} - y^{p^n} = x - y$ and thus $x - y \in \mathbb{F}$. If $x, y \in \mathbb{F} \setminus \{0\}$, then $(xy^{-1})^{p^n} = x^{p^n}(y^{p^n})^{-1} = xy^{-1}$ and thus $xy^{-1} \in \mathbb{F}$. So $\mathbb{F}$ is a field with p^n elements.

We will continue with the uniqueness. Assume $\mathbb{F}$ is a field of p^n elements. By Proposition 1.6.1 we can assume that $\mathbb{F}$ is an extension field of $\mathbb{Z}_p$, and thus $f \in \mathbb{F}[x]$. Since every element of $\mathbb{F}^*$ is a root of f by Little Fermat, one sees that $\mathbb{F}$ is the splitting field of f, and thus unique up to isomorphism. □

Notation 1.6.5. *Let $q = p^n$, where p is prime and $n \in \mathbb{N}_{>0}$. Then let $\mathbb{F}_q$ denote the finite field with q elements. Note that $\mathbb{F}_p \cong \mathbb{Z}_p$.*

From now on, let q be a prime power.

Proposition 1.6.6. *Let $\mathbb{F}$ be a finite field. Then the multiplicative group $\mathbb{F}^*$ is cyclic.*

Proof. It is clear that $\mathbb{F}^*$ is a finite Abelian group. By the structure theorem for finite Abelian groups, we have

$$\mathbb{F}^* \cong \mathbb{Z}_{n_1} \oplus \cdots \oplus \mathbb{Z}_{n_r},$$

where n_1 divides n_2, …, n_{r-1} divides n_r. Thus $\alpha^{n_r} = 1$ for all $\alpha \in \mathbb{F}^*$ by Little Fermat. Then all $\alpha \in \mathbb{F}^*$ are roots of $f := x^{n_r} - 1 \in \mathbb{F}[x]$, and thus $|\mathbb{F}^*| \leq n_r$, which implies $n_r = |\mathbb{F}^*|$ and thus $\mathbb{F}^* \cong \mathbb{Z}_{n_r}$. □

This implies $\mathbb{F}_q^* \cong \mathbb{Z}_{q-1}$ and $\mathbb{Z}_p^* \cong \mathbb{Z}_{p-1}$ (as groups!).

Remarks 1.6.7.

(1) *By using Proposition 1.6.6 we can show that there exists a unique field of p^n elements up to isomophism: If $\mathbb{F}$ and $\mathbb{K}$ are two such fields, there are $\alpha \in \mathbb{F}$ and $\beta \in \mathbb{K}$ such that $\mathbb{F}^* = \langle\alpha\rangle$ and $\mathbb{K}^* = \langle\beta\rangle$. Define the map $\psi : \mathbb{F} \to \mathbb{K}$ by $0 \mapsto 0$ and $\alpha^n \mapsto \beta^n$. It is not hard to show that this is an isomorphism.*

(2) *The proposition does* not *gives us a way to find the generators of $\mathbb{F}^*$.*

(3) *Consider the* Discrete Logarithm Problem (DLP)*:*

Let $\mathbb{F}^* = \langle\alpha\rangle$ and $\beta \in \mathbb{F}^*$. Can we find some n such that $\alpha^n = \beta$? (I. e. $n = \log_\alpha \beta$.)

This problem is very hard if $|\mathbb{F}|$ is "big", and β is "general" (i. e. chosen at random).

Corollary 1.6.8. *Every finite field $\mathbb{F}$ can be represented as $\mathbb{F} \cong \mathbb{Z}_p[x]/(f)$ where $f \in \mathbb{Z}_p[x]$ is irreducible. If $\operatorname{Char}\mathbb{F} = p$ and $\deg f = n$, then $|\mathbb{F}| = p^n$.*

Proof. Define a ring homomorphism

$$\psi : \mathbb{Z}_p[x] \to \mathbb{F}, \qquad 1 \mapsto 1, \qquad x \mapsto \alpha,$$

where $p = \operatorname{Char}\mathbb{F}$ and $\alpha \in \mathbb{F}^*$ generates $\mathbb{F}^*$ as a group. This map is surjective, and thus $\mathbb{F} \cong \mathbb{Z}_p[x]/\ker\psi$. Now $\mathbb{Z}_p[x]$ is a principle ideal domain (PID) and thus $\ker\psi = (f)$ for an $f \in \mathbb{Z}_p[x]$. Since $\mathbb{F}$ is a field, (f) must be maximal and thus f irreducible. Since

$$|\mathbb{Z}_p[x]/(f)| = p^{\deg f}$$

we conclude. □

Corollary 1.6.9. *There exists at least one irreducible polynomial of degree $n \in \mathbb{N}_{>0}$ in $\mathbb{Z}_p[x]$ for all primes p.*

Proof. Represent $\mathbb{F}_{p^n}$ by $\mathbb{Z}_p[x]/(f)$ as in the last corollary; then f is irreducible of degree n. □

Remark 1.6.10.

(1) *If $f \in \mathbb{Z}_p[x]$ is the minimal polynomial of a generator as in the proof of the lemma, and $\mathbb{F} \cong \mathbb{Z}_p[x]/(f)$, then $\overline{x}$ is a generator of $\mathbb{F}^*$.*

(2) *Let $\mathbb{F}^* = \langle\alpha\rangle$ and $|\mathbb{F}| = p^n$. Take $1, \alpha, \ldots, \alpha^n \in \mathbb{F}$. Since $\dim_{\mathbb{Z}_p}\mathbb{F} = n$, these elements are linearly dependent and thus there exist $a_0, \ldots, a_n \in \mathbb{Z}_p$, not all zero, such that*

$$\sum_{i=0}^{n} a_i\alpha^i = 0.$$

Let $f = \sum a_i x^i \in \mathbb{Z}_p[x]$. Then f is the minimal polynomial[3] of x over $\mathbb{Z}_p$, and $\mathbb{F} \cong \mathbb{Z}_p[x]/(f)$.

Theorem 1.6.11. *The multiplicative group $\mathbb{F}^*_{p^n}$ embeds in a natural way in $GL_n(\mathbb{Z}_p)$.*

Proof. For $n = 1$ this is clear, so let $n > 1$. Define the ring morphism

$$\varphi : \mathbb{F}_{p^n} \cong \mathbb{Z}_p[x]/(f) \to \mathbb{Z}_p^{n\times n}, \qquad 0 \mapsto 0, \qquad \overline{x} \mapsto A$$

where

$$A = \begin{pmatrix} 0 & & 0 & -a_0 \\ 1 & \ddots & & \vdots \\ & \ddots & 0 & (-1)^{n-1}a_{n-2} \\ 0 & & 1 & (-1)^n a_{n-1} \end{pmatrix} \qquad \text{such that} \quad x^n + \sum_{i=0}^{n-1} a_i x^i = f.$$

This is well-defined: It is easy to see that $\det A = (-1)^{n+1}a_0 \neq 0$ since f is irreducible, and thus $\psi|_{\mathbb{F}^*_{p^n}}$ is a well defined map from $\mathbb{F}^*_{p^n}$ to $GL_n(\mathbb{Z}_p)$. Further $\det(\lambda I_n - A) = f(\lambda)$, and thus

$$\psi(\overline{f}) = f(\psi(\overline{x})) = f(A) = 0$$

by Cayley-Hamilton, since f is the characteristic polynomial of A.

The map ψ is injective, since $\psi(\overline{g(x)}) = 0$ implies $g(A) = 0$ and thus $f \mid g$, since f is also the minimal polynomial of A because it is irreducible. □

[3]This means that f is monic, i. e. the highest coefficient is one, and minimal in the sense that if $g \in \mathbb{Z}_p[x]$ is another polynomial vanishing at α, then f divides g.

Remarks 1.6.12. *In the exercises we found out that*

$$|GL_n(\mathbb{Z}_p)| = \prod_{i=0}^{n-1}(p^n - p^i),$$

and further $\left|\mathbb{F}_{p^n}^*\right| = p^n - 1$. *Thus if* $n > 1$, *then* $|GL_n(\mathbb{Z}_p)|$ *is larger than* $\left|\mathbb{F}_{p^n}^*\right|$. *For* $n = 1$, *we have* $\mathbb{F}_p^* \cong \mathbb{Z}_p^* \cong GL_1(\mathbb{Z}_p)$.

Definition 1.6.13. *The* Galois group $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$ *of* $\mathbb{F}_{p^n}$ *over* $\mathbb{F}_p$ *is*

$$\{\varphi : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n} \mid \varphi \text{ ring homomorphism where } \varphi|_{\mathbb{F}_p} = \mathbf{id}_{\mathbb{F}_p}\}.$$

If $\varphi \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$, it must be that $\ker\varphi \in \{0, \mathbb{F}_{p^n}\}$ since $\mathbb{F}_{p^n}$ is a field. Since $\varphi(1) = 1$ we get that φ is injective, and since $\mathbb{F}_{p^n}$ is finite, φ must also be surjective. So φ is an automorphism of $\mathbb{F}_{p^n}$ and it is easy to verify that $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$ is a group.

Examples 1.6.14.

1. *It is* $\mathrm{Gal}(\mathbb{F}_p/\mathbb{F}_p) = \{\mathbf{id}\}$.
2. *Define* $F : \mathbb{F}_{p^n} \to \mathbb{F}_{p^n}$ *by* $x \mapsto x^p$. *Then* $F|_{\mathbb{F}_p}$ *is the identity on* $\mathbb{F}_p$ *by Little Fermat, and thus* $F \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$. *We call* F *the* Frobenius endomorphism. *Note that* $\langle F\rangle = \{\mathbf{id}, F, F^2, F^3, \dots\}$ *is a subgroup of* $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$.

Theorem 1.6.15. *The Frobenius endomorphism* F *generates* $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$.

Proof. Let $\varphi \in \mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$ and let $\alpha \in \mathbb{F}_{p^n}^*$ such that $\langle\alpha\rangle = \mathbb{F}_{p^n}^*$. Let $k \in \{1, \dots, p^n - 1\}$ such that $\alpha^k = \varphi(\alpha)$. Assume $k > 1$, since otherwise $\varphi = \mathbf{id} = F^0$. Then $\varphi(x) = x^k$ for every $x \in \mathbb{F}_{p^n}$ since φ is a ring homomorphism and $\mathbb{F}_{p^n} = \{0\} \cup \langle\alpha\rangle$.

Write $k = p^\ell r$ where p does not divides r. Then $\varphi \circ F^{n-\ell}$ maps every x onto $(x^{p^{n-\ell}})^{p^\ell r} = x^{p^n r}$, and since $x^{p^n} - x$ annihilates every element of $\mathbb{F}_{p^n}$ we have that $x^{p^n r}$ is the same than x^r. We now want to show $r = 1$. Without loss of generality we can assume $k = r$, i. e. p does not divides k. Assume that $k > 1$.

Take a look at the polynomial $f := (x+1)^k - x^k - 1 = \sum_{i=1}^{k-1}\binom{k}{i}x^i \in \mathbb{F}_{p^n}[x]$. This polynomial is annihilated by every element of $\mathbb{F}_{p^n}$, since $x \mapsto x^k$ is a ring endomorphism of $\mathbb{F}_{p^n}$. Thus $\deg f$ must be at least p^n, or $f = 0$. Since $r < 0$ this means that p divides $\binom{k}{i}$ for every $i = 1, \dots, p-1$, and especially $\binom{k}{1} = k$. But this is a contradiction! □

What Galois theory says is that there is a one-to-one correspondence between subfields of $\mathbb{F}_{p^n}$ which contain $\mathbb{F}_p$ and subgroups of $\mathrm{Gal}(\mathbb{F}_{p^n}/\mathbb{F}_p)$. A subfield corresponds to the subgroup which leaves the subfield fixed. And a subgroup corresponds to the subfield which is left fixed by every element of the subgroup.

Let m be a divisor of n. Then the elements of $\mathbb{F}_{p^n}$ which are fixed under F^m are exactly the elements of $\mathbb{F}_{p^m}$, since $F^m(\alpha) = \alpha$ if and only if α is a root of $x^{p^m} - x$, and $\mathbb{F}_{p^m}$ is the splitting field of $x^{p^m} - x$.

If $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$, then $\mathbb{F}_{p^n}$ is an $\mathbb{F}_{p^m}$-vector space and thus p^n is a power of p^m, which implies that m divides n. Thus we have shown that $\mathbb{F}_{p^m}$ is a subfield of $\mathbb{F}_{p^n}$ if and only if m divides n.

1.7 Security Issues of RSA

Recall that $n = pq$, where $p, q \geq 10^{100}$ are prime. The public information are the modulus n, the encryption exponent e and the encryption map $\psi : \mathbb{Z}_n \to \mathbb{Z}_n$, $m \mapsto m^e = c$. The private information are the primes p and q and the decryption exponent d, where $ed \equiv 1 \pmod{\phi(n)}$. Further, decryption is done by $\psi^{-1} : \mathbb{Z}_n \to \mathbb{Z}_n$, $c \mapsto c^d = m$.

The fundamental question is: is being able to break RSA (that is computing ψ^{-1}) polynomial equivalent to factoring n?

Lemma 1.7.1. *Knowing p and q is polynomial equivalent to knowing n and $\phi(n)$.*

Proof. Consider the relations $n = pq$ and $\phi(n) = (p-1)(q-1)$. If n and $\phi(n)$ are known, one can find p and q by solving this quadratic equation over the reals. The other direction is trivial. □

Lemma 1.7.2. *Knowing the decryption exponent d is polynomial equivalent to factor.*

Proof. If p, q and e are known, d can easily be computed. The other direction is more involved; we only give an outline of the proof.

Given d, it follows that $m^{de-1} \equiv 1 \pmod{n}$ for all $m \in \mathbb{Z}_n^*$. It follows that $\phi(n)$ divides $de - 1$. Let $k = de - 1$ and write $k = 2^t r$ with r odd. Since p and q are odd, $\phi(n)$ is divisible at least by four and thus $r \geq 2$.

Let $g \in \mathbb{Z}_n^*$ be randomly chosen. Consider the sequence

$$g^r,\ g^{2r},\ \ldots,\ g^{2^t r}.$$

Let i be the smallest index such that $g^{2^i r} \equiv 1 \pmod{n}$. Then $g^{2^{i-1} r} \not\equiv 1 \pmod{n}$ if $i \geq 1$.

By the Chinese Remainder Theorem $\mathbb{Z}_n \cong \mathbb{Z}_p \times \mathbb{Z}_q$, and thus $g^{2^{i-1} r}$ maps to $(\pm 1, \pm 1)$. So there are four possibilities for $g^{2^{i-1} r}$:

(a) it corresponds to $(1, 1)$; thus $g^{2^{i-1} r} \equiv 1 \pmod{n}$; this will not happen by hypothesis;

(b) it corresponds to $(-1, -1)$; thus $g^{2^{i-1} r} \equiv -1 \pmod{n}$;

(c) it corresponds to $(1, -1)$ or $(-1, 1)$, and thus $\gcd(g^{2^{i-1} r}, n)$ is either p or q.

One can show that for randomly chosen g, more than fifty percent of the cases one deals with are case (c). The proof for this is left to the reader as an exercise.

The cost of this algorithm is $\mathcal{O}(\log^3 n)$. □

1.7.1 Implementation Weaknesses

(1) **p and q should be sufficiently apart:** For example, the following is a bad choice: let a be a random number around 10^{100}. Let $p := nextprime(a)$ and $q := nextprime(p + 1)$, and $n := pq$. This can be attacked since $q = nextprime(\sqrt{n})$.

(2) **Pollards $(p - 1)$ factoring attack:**

Definition 1.7.3. *Let m and B be positive integers. One says that m is B-*smooth *if all prime factors of m are less or equal than B.*

Example 1.7.4. *The number 96 is 3-smooth: it is $96 = 2^5 \cdot 3$.*

Assume $n = pq$ and that $p - 1$ is B-smooth, but $q - 1$ is not (for a small bound B). Define

$$k := \prod_{\substack{\alpha \leq B \\ \alpha \text{ prime}}} \alpha^{\left\lfloor \frac{\log n}{\log \alpha} \right\rfloor}.$$

By assumption $q - 1$ does not divides k, but $p - 1$ does. By little Fermat we have

$$a^k \equiv 1 \pmod{p} \qquad \text{and} \qquad a^k \not\equiv 1 \pmod{q}$$

for more than fifty percent of the a's. (Another exercise for the interested reader.) If $a^k \not\equiv 1 \pmod{q}$, then $\gcd(a^k - 1, n) = p$.

Remark 1.7.5. *For randomly chosen p, with a high probability $p-1$ has a large prime factor.*

Definition 1.7.6. *An odd prime p is called a* safe prime *if $\frac{p-1}{2}$ is prime.*

Examples 1.7.7. *The numbers 7 and 11 are safe primes.*

In practice, p and q are chosen as safe primes.

(3) **Common modulus attack:** A sitation: A large corporation computes $n = pq$ with p, q safe primes. Different web servers get pairs (e_i, d_i) of encryption/decryption exponents for this modulus n. As p and q are safely stored (maybe even decentralized), the compromise of one server does not compromises the others.

But this assumption is wrong, as by one of the above lemmata p and q can be computed from one pair (e_i, d_i).

In addition, if the same modulus is used with two different encryption exponents e_1 and e_2 which are coprime, and a message can be intercepted both encrypted by e_1 and e_2, then the original message can be decrypted without breaking the system itself. (See the exercises.)

(4) **Short message encryption:** In practice n is around 1024 bits. Assume a message $1 \leq m \leq 2^{40}$ is send.

With probability around 18 percent, $m = m_1 m_2$ with $m_1, m_2 \leq 2^{22}$. Then $c \equiv m^e \equiv m_1^e m_2^e \pmod{n}$. Produce a list of $\frac{c}{m_1^e} \mod n$ for $1 \leq m_1 \leq 2^{22}$ and store the last 50 bits of each result. Compute $m_2^e \mod n$ for $1 \leq m_2 \leq 2^{22}$ and check if the last 50 bits agree with a number in the previous list. This leaves a short list of canidates for $\frac{c}{m_1^e} \equiv m_2^e \pmod{n}$; in that case we found $m = m_1 m_2$.

(5) **Bleichenbacher attack (1998):** Under public key cryptography standard PKCS I, n is chosen to have 1024 bits, and the following protocol is used: Of each message m, the first 16 bits specify the protocol ID, then there follow a lot of random bits, followed by some zeros to indicate the start of the real message, and then the last 128 bits contain the real message.

The default behaviour for a server who received such a packet which contained an invalid protocol ID was to send the invalid protocol ID back to the sender, in decrypted form.

Bleichenbacher exploited this behaviour to produce a decryption of $c = m^e \mod n$ bit-by-bit by sending many (invalid) requests to the server, which are of the form

$$c' = cr^e = (mr)^e \mod n.$$

As an exercise, figure out how this can be done. Hint: Multiplying by two is (more or less) a cyclic shift.

(6) **Low public key:** Early implementations used $e = 3$ as an encryption exponent. (This has the advantage that only two multiplications modulo n are needed for encryption.) There are several attacks known; the most sophisticated is by Coppersmith using shortest vector computation with LLL.

Another reason: If $n \approx 2^{1024}$ and $m \leq 2^{300}$, then $m^3 \leq 2^{900}$ and thus $m^3 \mod n = m^3 \in \mathbb{N}$. So by taking the cubic root of $m^3 \mod n$ over $\mathbb{R}$ gives m.

A third attack is the following: Assume m^3 is known for different moduli, for example $m^3 \equiv c_i \pmod{n_i}$, where $i = 1, \ldots, 4$. Without loss of generality $\gcd(n_i, n_j) = 1$ for $i \neq j$. Under reasonable assumptions we can expect $m^3 < \prod_{i=1}^4 n_i =: n$. By the Chinese Remainder Theorem, we can reconstruct $m^3 \mod n$ and thus $m^3 \in \mathbb{N}$ from $m^3 \pmod{n_i}$. Thus again we can take the cubic root in $\mathbb{R}$ to get m.

In practice, it is better to use $e = 2^{16} + 1 = 65537$; this is also prime and fairly easy to compute.

(7) **Low private key exponent:** Another tempting idea is to let $d = 3$; then for example a web server's load is reduced dramatically. But there are several reasons why this is bad:

First, a too small d is bad since just trying $d = 2, 3, \ldots$ (small numbers) gives back m from m^e.

Second, in 1990 M. Wiener shows that d should be at least $n^{1/4}$. The idea of Wiener is that $ed - b\phi(n) = 1$ for some $b \in \mathbb{Z}$, and thus

$$\left|\frac{e}{\phi(n)} - \frac{b}{d}\right| = \frac{1}{d\phi(n)}.$$

Assume $1 \leq p < q \leq 2p$ and $d < n^{1/4}$. Then

$$|n - \phi(n)| \leq 3\sqrt{n}, \qquad \text{so} \qquad \left|\frac{e}{n} - \frac{b}{d}\right| \leq \frac{1}{dn^{1/4}} < \frac{1}{2d^2}.$$

By using continued fraction expansion, b and d can be found (or at least a short list of canidates).

!!! ??? $n^{1/2}$ anstelle $n^{1/4}$ in der Formel, da spaeter d^2 ??? !!!

A **conclusion**: in an implemtation all difficulties above are taken into account nowadays. The security depends mainly on the difficulty of factoring.

1.7.2 Some Quick Notes on Factoring

A major idea in factoring is the "quadratic sieve": Consider the polynomial $f := x^2 - y^2 \in \mathbb{Z}[x, y]$. Assume $(\alpha, \beta) \in \mathbb{Z}^2$ is a point with $f(\alpha, \beta) = 0$. More generally, assume $f(\alpha, \beta) \equiv 0 \pmod{n}$, where n is the product of two distinct primes p and q. There are four possibilities:

(a) It is $\alpha \equiv \beta \pmod{n}$;

(b) It is $\alpha \equiv -\beta \pmod{n}$;

(c) p divides $\alpha + \beta$ and q divides $\alpha - \beta$;

(d) q divides $\alpha + \beta$ and p divides $\alpha - \beta$,

In cases (c) and (d), computing $\gcd(\alpha - \beta, n)$ reveals a factor of n. But how to get a non-trivial solution of $f(\alpha, \beta) \equiv 0 \pmod{n}$?

First, chose a factor base $p_1, \ldots, p_m$ (distinct primes); for example the first m primes. Then search for numbers $x_i \in \mathbb{Z}_n$ such that $x_i^2 \mod n$ can be completely factored over $p_1, \ldots, p_m$. Write

$$x_i^2 \mod n = p_1^{e_{1i}} \cdots p_m^{e_{mi}}, \qquad i = 1, \ldots, \ell.$$

From this we produce a binary matrix $(e_{ij} \mod 2)_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq m}}$. Find $\lambda_i \in \mathbb{F}_2$ not all zero such that $\sum_{i=1}^{\ell} \lambda_i e_{ij}$ is even for all j (this is basic linear algebra over $\mathbb{F}_2$!). Then we found

$$x^2 = \prod_{i=1}^{\ell} (x_i^2)^{\lambda_i}$$

which is a square.

By using this, the RSA challenge 512 was solved at the end of the 90'th.

1.8 Secret Key Ciphers

Recall that a secret key cipher consists of two maps

$$\varphi : \mathcal{M} \times \mathcal{K} \to \mathcal{C}, \qquad \psi : \mathcal{C} \times \mathcal{K} \to \mathcal{M}$$

such that

(1) we have $\psi(\varphi(m,k),k) = m$ for every $k \in \mathcal{K}$ and $m \in \mathcal{M}$ and that

(2) for fixed $m \in \mathcal{M}$, the function $\varphi_m : \mathcal{K} \to \mathcal{C}$, $k \mapsto \varphi(m,k)$ is a one-way function.

In practice, there are two systems available:

(A) stream ciphers ($\mathcal{M}, \mathcal{K}, \mathcal{C}$ can have arbitrary sizes) and

(B) block ciphers ($\mathcal{M}, \mathcal{K}, \mathcal{C}$ are fixed finite sets).

1.8.1 Stream Ciphers

In 1917, Vernam invented (and got a patent) for the one-time pad. For this let $m = (m_0, m_1, m_2, \dots) \in \mathbb{Z}_2^{\mathbb{N}}$ (that are the $\mathbb{Z}_2$-valued sequences). Alice and Bob exchange a key $k = (k_i)_i \in \mathbb{Z}_2^{\mathbb{N}}$. The encryption is done by

$$c = m + k = (c_i + k_i)_i \in \mathbb{Z}_2^{\mathbb{N}},$$

and decryption by

$$m = c + k = (c_i + k_i)_i \in \mathbb{Z}_2^{\mathbb{N}}.$$

(This works since $c_i + k_i = m_i + 2k_i = m_i$ in $\mathbb{Z}_2$.)

In 1949, Shannon proved that the one-time pad is unconditionally and provable secure. In order to make this precise, Shannon viewed the sequences $(m_i)_i$, $(k_i)_i$ and $(c_i)_i$ as generated by random variables M, K and C. For a discrete[4] random variable X he introduced the notion of *entropy*:

$$H(X) := -\sum_{i=1}^{t} p_i \log_2 p_i, \qquad \text{where } P(X \in \{m_1, \dots, m_t\}) = 1,\ p_i = P(X = m_i)$$

and the m_i are pairwise distinct.

Examples 1.8.1.

(1) *Let X describe a Bernoulli trial with $p = \frac{1}{2}$ and $q = 1 - p = \frac{1}{2}$, i. e. $P(X = 1) = p$ and $P(X = 0) = q$. Then*

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2} = 1.$$

(2) *Let A, C, T and G have the probabilites $P(X = A) = P(X = C) = P(X = T) = P(X = G) = \frac{1}{4}$. In this case we have*

$$H(X) = -4 \cdot \log_2\frac{1}{4} = 2.$$

An encryption scheme would be

$$A \mapsto 00, \quad C \mapsto 01, \quad T \mapsto 10, \quad G \mapsto 11.$$

(3) *Now assume $P(X = A) = \frac{1}{2}$, $P(X = C) = \frac{1}{4}$ and $P(X = T) = P(X = G) = \frac{1}{8}$. What about this scheme:*

$$A \mapsto 0, \quad C \mapsto 10, \quad T \mapsto 110, \quad G \mapsto 111.$$

One can easily check that a sequence consisting of A, C, T and G encoded by this scheme can be uniquely decoded. How many bits per letter are needed in average? We have

$$E(\ell(X)) = \frac{1}{2}\cdot 1 + \frac{1}{4}\cdot 2 + \frac{1}{8}\cdot 3 + \frac{1}{8}\cdot 3 = \frac{7}{4} = 1.75 < 2.$$

The entropy is

$$H(X) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{8}\log_2\frac{1}{8} - \frac{1}{8}\log_2\frac{1}{8} = \frac{7}{4}.$$

[4] A random variable X is called discrete if there exists a finite set S such that $P(X \in S) = 1$.

The *Noiseless Shannon Theorem* says:

> Any encoding scheme of a random sample from a random variable X requires *at least* $H(X)$ bits per symbol. There are encoding schemes which in the limit can reach that bound.

In cryptography Shannon defined a secret key cryptosystem to be *unconditionally and provable secure* if

$$H(M \mid C) = H(M).$$

Here $H(M \mid C)$ denotes the conditionally entropy under the knowledge of the cipher of M.

He proved that the one-time pad is unconditionally secure as soon as $H(K) \geq H(M)$. The problem with this result is that a secret key has to be exchanged which is longer than the longest message ever sent.

The idea of stream ciphers is to generate pseudo-random sequences $(k_i)_i$ starting only with some finite data.

Example 1.8.2. *Let $s_{i+2} = s_i + s_{i+1}$ where $s_0 = s_1 = 1$; this is a Fibonacci sequence. Over $\mathbb{Z}$ the sequence looks like*

$$1, 1, 2, 3, 5, 8, \dots,$$

(i. e. there is no period), while modulo 3 (i. e. over $\mathbb{Z}_3$) it looks like

$$1, 1, 2, 0, 2, 2, 1, 0, 1, 1, \dots,$$

i. e. it has a period of 8. Modulo 16 it looks like

$$1, 1, 2, 3, 5, 8, 13, 5, 2, 7, 9, 0, 9, 9, 2, \dots,$$

so the period is quite long.

Definition 1.8.3. *Let $\mathbb{F}$ be a finite field. The relation*

$$s_{n+i} + b_{n-1}s_{n-1+i} + b_{n-2}s_{n-2+i} + \cdots + b_0 s_i = 0, \qquad (*)$$

where $s_i, b_j \in \mathbb{F}$ for $j = 0, \dots, n-1$ and $i \in \mathbb{N}$, is called an n-th order linear recurrence relation *having the* characteristic polynomial

$$\chi(z) = z^n + b_{n-1}z^{n-1} + \cdots + b_0 \in \mathbb{F}[z].$$

Example 1.8.4. *The Fibonacci sequence is given by the second order recurrence relation, whose characteristic polynomial is*

$$\chi(z) = z^2 - z - 1.$$

In the next paragraphs we want to inspect the following questions:

(1) How can the total solution space of $(*)$ be described?

(2) How can sequences with long periods be constructed?

(The largest possible period is $q^n - 1$, where $q = |\mathbb{F}|$.)

Let $V = \mathbb{F}^{\mathbb{N}} = \{(s_0, s_1, \dots) \mid s_i \in \mathbb{F}\}$ be the vector space of the $\mathbb{F}$-valued sequences. (Note that this is an infinite dimensional vector space with an uncountable basis.) Define the *shift map*

$$D : V \to V, \qquad (s_i)_i \mapsto (s_{i+1})_i.$$

This is an $\mathbb{F}$-linear map.

Lemma 1.8.5. *The element $s \in V$ satisfies $(*)$ if and only if $\chi(D)(s) = 0$, and that happens if and only if $s \in \ker \chi(D)$.*

Proof. Clear. □

Example 1.8.6. *For the Fibonacci sequence, we have* $\chi(z) = z^2 - z - 1$ *and*

$$(D^2 - D - 1)((s_i)_i) = (s_2 - s_1 - s_0, s_3 - s_2 - s_1, s_4 - s_3 - s_2, \dots).$$

The consequence is that the total solution space of $(*)$ is a subspace of V.

Lemma 1.8.7. *The dimension of* $\ker \chi(D)$ *over* $\mathbb{F}$ *is* $\deg \chi$.

Proof. This is also clear, since any choice of n initial conditions $s_i = \bar{s}_i$ for $i = 0, \dots, n-1$ determines a unique solution of $(*)$. □

Lemma 1.8.8. *Assume* $\chi_1, \chi_2 \in \mathbb{F}[z]$ *are given. Then we have*

$$\ker \chi_1(D) \subseteq \ker \chi_2(D) \iff \chi_1 \text{ divides } \chi_2.$$

Proof. If χ_1 divides χ_2, then $\chi_2 = r \cdot \chi_1$ where $r \in \mathbb{F}[x]$. If we have $w \in \ker \chi_1$, then $\chi_2(D)(w) = (r(D)\chi_1(D))(w) = r(D)(\chi_1(D)(w)) = r(D)(0) = 0$, thus we have $w \in \ker \chi_2(D)$.

For the other direction assume that $\ker \chi_1(D) \subseteq \ker \chi_2(D)$. Write $\chi_2 = q \cdot \chi_1 + r$ where $q, r \in \mathbb{F}[x]$ and $r = 0$ or $\deg r < \deg \chi_1$. Now for every $w \in \ker \chi_1(D)$ we have $r(D)(w) = \chi_2(D)(w) - q(D)(\chi_1(D)(w)) = 0$, thus $\ker \chi_1(D) \subseteq \ker r(D)$. Assume that $r \neq 0$: this implies that $\dim \ker r(D) \geq \dim \ker \chi_1(D) = \deg \chi_1$, but by the previous lemma $\dim \ker r(D) = \deg r < \deg \chi_1$, contradiction! □

Lemma 1.8.9. *Assume*

$$\chi(z) = \prod_{i=1}^{n} (z - \lambda_i)$$

where the $\lambda_i \in \mathbb{F}$ *are pairwise distinct. Then*

$$\ker \chi(D) = \ker(D - \lambda_1) \oplus \cdots \ker(D - \lambda_n).$$

Moreover for $\lambda \in \mathbb{F}$ *we have*

$$\ker(D - \lambda) = \mathbb{F} \cdot (\lambda^i)_i = \{(c, c\lambda, c\lambda^2, \dots) \mid c \in \mathbb{F}\}.$$

Proof. The form of $\ker(D - \lambda)$ is clear. By the last lemma we know $\ker(D - \lambda_i) \subseteq \ker \chi(D)$. Since the $(\lambda_i^j)_j$ are linearily independent[5] the claim follows. □

Remark 1.8.10. *When there are multiple roots, for example if* λ *is an* m*-th root of* χ*, then* $\ker(D - \lambda)^m$ *consists of*

$$\operatorname{span}\left\{(\lambda^i)_i, (i\lambda^{i-1})_i, \dots, (0, \dots, 0, \tfrac{(m-1)!}{0!}\lambda^0, \tfrac{m!}{1!}\lambda^1, \tfrac{(m+1)!}{2!}\lambda^2, \tfrac{(m+2)!}{3!}\lambda^3, \dots)\right\}.$$

(Note the similarity to homogenous linear differential equations: the other solutions for the root λ *are found by differentiating the first one.)*

Example 1.8.11. *Find an explicit formula for the Fibonacci sequence over* $\mathbb{F}_{19}$*. We have* $\chi(z) = z^2 - z - 1 = (z - 5)(z - 15)$*. The general solution is thus*

$$s_i = c_0 5^i + c_1 15^i, \qquad c_0, c_1 \in \mathbb{F}.$$

To get a particular sequence where $s_0 = s_1 = 1$*, we solve*

$$1 \overset{!}{=} s_0 = c_0 + c_1 \qquad \text{and} \qquad 1 \overset{!}{=} s_1 = 5c_0 + 15c_1,$$

leading to $c_0 = 9$ *and* $c_1 = 11$*. Thus*

$$s_i = 9 \cdot 5^i + 11 \cdot 15^i, \quad i \in \mathbb{N}$$

is the Fibonacci sequence over $\mathbb{F}_{19}$*!*

[5] Which follows directly from writing λ_i^j where $1 \leq i \leq n$, $0 \leq j < n$ into a matrix. This matrix has a special form and is called a *Vandermonde matrix*. The determinant of this is nonzero if and only if the λ_i are pairwise distinct, which we have required here.

Another way is to use *generating functions*: We define

$$\mathbb{F}((z)) = \left\{ \sum_{i=-N}^{\infty} a_i z^i \mid N \in \mathbb{N},\ a_i \in \mathbb{F} \right\} = \mathbb{F}[[z]] \oplus z^{-1}\mathbb{F}[z^{-1}]$$

to be the ring of *formal Laurent series in* z. (Note that $\mathbb{F}((z))$ is the quotient field of $\mathbb{F}[[z]]$, the *formal power series* in z.) In the following, we need the formal Laurent series in z^{-1},

$$\mathbb{F}((z^{-1})) = z^{-1}\mathbb{F}[[z^{-1}]] \oplus \mathbb{F}[z].$$

Consider the vector space

$$V = z^{-1}\mathbb{F}[[z^{-1}]] = \mathbb{F}((z^{-1}))/\mathbb{F}[z] \mathrel{\hat{=}} \left\{ \tfrac{s_0}{z} + \tfrac{s_1}{z^2} + \tfrac{s_2}{z^3} + \cdots \mid s_i \in \mathbb{F} \right\};$$

this space is isomorphic to $\mathbb{F}^{\mathbb{N}}$.

Remark 1.8.12. *Multiplication by z inside V corresponds to the shift map D in $\mathbb{F}^{\mathbb{N}}$!*

Lemma 1.8.13. *Let $s = (s_i)_i \in \mathbb{F}^{\mathbb{N}}$ and $f(z) = \frac{s_0}{z} + \frac{s_1}{z^2} + \frac{s_2}{z^3} + \cdots \in V$. Then s satiesfies $(*)$ if and only if $f(z) = \frac{r(z)}{\chi(z)}$ with $r \in \mathbb{F}[z]$ such that $\deg r < \deg \chi$.*

Proof. We have $\chi(D)(s) = 0$ if and only if $\chi(z)f(z) = r(z) \in \mathbb{F}[z]$. □

Example 1.8.14. *Again the Fibonacci sequence: We want to find an explicit formula for the Fibonacci sequence $s_{i+2} = s_{i-1} + s_i$ where $s_0 = s_1 = 1$ using generating functions. The general solution of $s_{i+2} - s_{i+1} - s_i = 0$ (in the sense of the previous lemma) is given by*

$$f(z) = \frac{a_1 z + a_0}{z^2 - z - 1} = \sum_{i=0}^{\infty} \frac{s_i}{z^{i+1}}.$$

The initial condition $s_0 = s_1 = 1$ results in $a_1 = 1$ and $a_0 = 0$, since we get

$$f(z) = \frac{z}{z^2 - z - 1} = \frac{z}{z^2} \cdot \frac{1}{1 - (\frac{1}{z} + \frac{1}{z^2})} = \frac{1}{z} + \frac{1}{z^2} + \frac{1}{z^3} + \cdots.$$

Let $z^2 - z - 1 = (z - \alpha_1)(z - \alpha_2)$. By partial fraction composition, we get

$$f(z) = \frac{z}{z^2 - z - 1} = \frac{A}{z - \alpha_1} + \frac{B}{z - \alpha_2} \qquad \text{where } A = \frac{\alpha_1}{\alpha_1 - \alpha_2} \text{ and } B = \frac{\alpha_2}{\alpha_2 - \alpha_1}.$$

If $\mathbb{F} = \mathbb{C}$ we have $\alpha_1 = \frac{1}{2}(1 + \sqrt{5})$ and $\alpha_2 = \frac{1}{2}(1 - \sqrt{5})$, and by using the geometric series we have

$$\frac{1}{z - \beta} = \frac{1}{z} \cdot \frac{1}{1 - \frac{\beta}{z}} = \frac{1}{z} \sum_{i=0}^{\infty} \frac{\beta^i}{z^i} = \sum_{i=1}^{\infty} \frac{\beta^{i-1}}{z^i}.$$

Thus we get

$$\frac{z}{z^2 - z - 1} = \frac{1}{\sqrt{5}} \sum_{i=1}^{\infty} \frac{(1 + \sqrt{5})^i}{2^i z^i} - \frac{1}{\sqrt{5}} \sum_{i=1}^{\infty} \frac{(1 - \sqrt{5})^i}{2^i z^i},$$

and so we have

$$s_i = \frac{1}{\sqrt{5}} \left(\frac{1 + \sqrt{5}}{2} \right)^{i+1} - \frac{1}{\sqrt{5}} \left(\frac{1 - \sqrt{5}}{2} \right)^{i+1}.$$

In complex analysis is shown that any holomorphic (i. e. complex differentiable) function is of the form

$$f(z) = \sum_{i=0}^{\infty} a_i (z - z_i)^i, \qquad a_i, z, z_0 \in \mathbb{C}.$$

This leads to the question whether and how it is possible to determine if $f(z)$ is a rational function, i. e. $f(z) = \frac{g(z)}{h(z)}$ where $g, h \in \mathbb{C}[z]$.

Example 1.8.15. *Let* $f(z) = z + z^2 + 2z^3 + 3z^4 + 5z^5 + 8z^6 + 13z^7 + \dots$ *(Fibonacci coefficients) is rational, and*

$$f(z) = \frac{1/z}{(1/z)^2 - (1/z) - 1} = \frac{z^2}{1 - z - z^2}.$$

Definition 1.8.16. *A sequence* $s = (s_i)_i$ *is called* ultimately periodic *if there are numbers* $\hat{r}$, $\hat{j}$ *such that* $s_{\hat{r}+i} = s_i$ *for all* $i \geq \hat{j}$. *The smallest numbers* $\hat{r}$ *and* $\hat{j}$ *with the above properties are called the* period *and the* pre-period, *respectively.*

Example 1.8.17. *The sequence* $3, 7, 11, 5, 9, 2, 5, 2, 5, 2, 5, 2, 5, \dots$ *has period* 2 *and pre-period* 6.

Theorem 1.8.18 (Kronecker). *For a power series*

$$f(z) = \sum_{i=0}^{\infty} \frac{s_i}{z^{i+1}} \in \mathbb{F}[[z^{-1}]]$$

the following are equivalent:

(i) $f(z)$ *is a rational function of degree* n, *where the degree of* $f(z) = \frac{g(z)}{\chi(z)}$ *is defined*[6] *as* $\deg f := \max\{\deg g, \deg \chi\}$;

(ii) $s = (s_i)_i$ *satisfies the* n*-th order recurrence* $\chi(D)(s) = 0$;

(iii) *the Hankel (sp?) matrix*

$$H_f = \begin{pmatrix} s_0 & s_1 & s_2 & \cdots \\ s_1 & s_2 & & \\ s_2 & & \ddots & \\ \vdots & & & \end{pmatrix}$$

(an infinite matrix) has rank n.

If $\mathbb{F}$ *is finite, these are further equivalent to*

(iv) $s_0, s_1, s_2, \dots$ *is ultimately periodic.*

If $\mathbb{F}$ *is arbitrary and (iv) holds, this also implies (i)–(iii).*

Proof. The implication (ii) ⇒ (i) is the previous lemma: given $\chi(D)(s) = 0$, we have $f(z) = \frac{g(z)}{\chi(z)}$ where $\deg g < \deg \chi$.

(i) implies (ii): let $f(z) = \frac{g(z)}{\chi(z)} = \frac{\tilde{g}(z)}{\chi(z)} + r(z)$, where $r \in \mathbb{F}[z]$ and $\deg \tilde{g}(z) < \deg \chi$. Then the sequence $f(z) = \sum \frac{s_i}{z^{i+1}}$ satisfies $\chi(D)(s) = 0$.

(i) is equivalent to (iii): let

$$f(z) = \frac{\sum_{i=0}^{n-1} a_i z^i}{\sum_{i=0}^{n} b_i z^i} = \sum_{i=0}^{\infty} \frac{s_i}{z^{i+1}}.$$

This is equivalent to

$$\sum_{i=0}^{n-1} a_i z^i = \left(\sum_{i=0}^{n} b_i z^i\right)\left(\sum_{i=0}^{\infty} \frac{s_i}{z^{i+1}}\right),$$

and by comparing coefficients we get

$$\begin{aligned}
z^{n-1}: &\quad a_{n-1} = s_0, \\
z^{n-2}: &\quad a_{n-2} = b_{n-1}s_0 + s_1, \\
&\quad \vdots \qquad\qquad \vdots \\
z^0: &\quad a_0 = b_1 s_0 + \cdots + b_{n-1}s_{n-2} + s_{n-1}, \\
z^{-1}: &\quad 0 = b_0 s_0 + \cdots + b_{n-1}s_{n-1} + s_n, \\
z^{-k}: &\quad 0 = b_0 s_{k-1} + \cdots + b_{k-1}s_{n-k} + s_{n-k+1}, \qquad k \in \mathbb{N}.
\end{aligned}$$

[6] The rationale behind this definition comes from complex analysis: if $f(z) = \frac{g(z)}{h(z)}$ is reduced, i. e. g and h are coprime, then f is a d-to-one map from $\overline{\mathbb{C}}$ onto $\overline{\mathbb{C}}$, where $\overline{\mathbb{C}}$ is the Riemann sphere and $d = \deg f := \max\{\deg g, \deg h\}$.

This again is equivalent to

$$\begin{pmatrix} s_0 & s_1 & s_2 & \cdots \\ s_1 & s_2 & & \\ & & \ddots & \\ s_2 & & & \\ \vdots & & & \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-1} \\ \vdots \end{pmatrix} = 0,$$

which in turn is equivalent to that H_f has finite rank n.

(iv) implies (ii): assume $s = (s_i)_i$ is ultimately periodic with period r and pre-period j. Then $(D^{r+j} - D^j)(s) = 0$, and thus

$$f(z) = \frac{g(z)}{z^{r+j} - z^j},$$

where the fraction is not necessarily reduced.

(ii) implies (iv) in the case that $\mathbb{F}$ is finite: assume $\mathbb{F} = \mathbb{F}_q$ and $\chi(z) = z^n + \sum_{i=0}^{n-1} b_i z^i$, and $\chi(D)(s) = 0$. Introduce the *state vector* at time t,

$$x_t = \begin{pmatrix} s_{t+1} \\ \vdots \\ s_{t+n} \end{pmatrix},$$

and the *state transition matrix*

$$A = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -b_0 & \cdots & \cdots & -b_{n-1} \end{pmatrix}.$$

Then $(*)$ is equivalent to $x_{i+1} = Ax_i$ for all i. Consider the state sequence $x_0, x_1, x_2, \ldots$. Since $|\mathbb{F}^n| = |\mathbb{F}|^n = q^n < \infty$ by the pigeonhole principle[7] there exists $a, b \in \mathbb{N}$ with $0 \leq a < b \leq q^n$ such that $x_a = x_b$. Then $s = (s_i)_i$ is periodic of period at most $b - a \leq q^n$. □

Remark 1.8.19. *An addition to the implication (ii) ⇒ (iv): if a state is non-zero and $b_0 \neq 0$, then all states $x_0, x_1, x_2, \ldots$ are non-zero and the maximal period is thus strictly less than q^n.*

The question remains whether we can construct periodic sequences of period $q^n - 1$?

Lemma 1.8.20. *Let $\varphi(z) = z^n + \sum_{i=0}^{n-1} b_i z^i \in \mathbb{Z}_q[z]$ and $\varphi(0) = b_0 \neq 0$. Then there exists an $e \in \mathbb{N}$ such that $1 \leq e \leq q^n - 1$ and q divides the polynomial $z^e - 1$.*

Definition 1.8.21. *The smallest e with the property as in the lemma is called the* order *of φ.*

Proof of the lemma. Consider the ring $R = \mathbb{F}[z]/(\varphi)$ which has q^n elements. Consider the residue classes $z^i + (\varphi)$, where $i = 0, \ldots, q^{n-1}$. By the pigeonhole principle (note that $z^i \not\equiv 0 \pmod{\varphi}$ for all i) there exists r, s with $0 \leq r < s < q^n$ such that $z^s \equiv z^r \pmod{\varphi}$. Thus $z^r(z^{s-r} - 1) \equiv 0 \pmod{\varphi}$, and so φ divides $z^r(z^{s-r} - 1)$. Since $\varphi(0) \neq 0$ we get that φ divides $z^{s-r} - 1$. □

Lemma 1.8.22. *Let $\mathbb{F}_q$ be a finite field and let $\varphi(z)$ be irreducible. Then the order of φ is equal to the order of any of the roots of φ.*

Proof. If α is a root of φ, then the others are given by α, α^q, α^{q^2}, …, $\alpha^{q^{n-1}}$, and all of them have the same order ℓ. Thus φ divides $z^\ell - 1$, and ℓ is the smallest number with this property. □

Lemma 1.8.23. *Assume α is a generator of $\mathbb{F}_{q^n}^*$. Let $\varphi \in \mathbb{F}_q[z]$ be the minimal polynomial of α over $\mathbb{F}_q$. Then φ has order $q^n - 1$.*

[7] The *pigeonhole principle* states that if $n + 1$ objects are placed in n boxes, at least one box must contain two objects.

Proof. Clearly $\alpha^{q^n-1} = 1$ and thus φ divides $z^{q^n-1} - 1$. Since $\alpha^a \neq 1$ for $0 < a < q^n - 1$ we get that φ does not divides $z^a - 1$ for $0 \leq a < q^n - 1$. □

Corollary 1.8.24. *Let $\varphi = z^n + \sum_{i=0}^{n-1} b_i z^i$ be as above, and assume $s = (s_i)_i \in \mathbb{F}_q^{\mathbb{N}}$ is a sequence defined through $s_0 = \cdots = s_{n-1} = 0$ and*

$$s_{n+i} + b_{n-1}s_{n+i-1} + \cdots + b_0 s_i = 0 \qquad \text{for all } i \in \mathbb{N}$$

(that is, $\varphi(D)(s) = 0$). Then s is periodic with period $q^n - 1$ and pre-period 0.

Proof. Since $b_0 \neq 0$ it is possible to "reverse" the time direction, i. e. one can compute s_i from $s_{i+1}, \ldots, s_{i+n}$. This implies that there is no pre-period.

Since φ divides z^{q^n-1} it follows that $(D^{q^n-1} - 1)(s) = 0$, and thus s has a period dividing $q^n - 1$. But since φ does not divides $z^a - 1$ for $a < q^n - 1$, we get

$$(D^a - 1)(s) \neq 0 \qquad \text{for all } 0 < a < q^n - 1,$$

and thus the period is $q^n - 1$. □

But now we want to return to cryptography. Between 1940 and 1970 secret key ciphers were constructed where Alice and Bob agree on a minimal polynomial φ as above. They compute $f(z) = \frac{1}{\varphi(z)} = \sum_{i=0}^{\infty} \frac{s_i}{z^{i+1}}$. If Alice wants to send the message

$$m = \sum_{i=0}^{\infty} \frac{m_i}{z^{i+1}},$$

then she computes $c = f + m$ and sends c. Bob deciphers it via $m = c - f$.

Remark 1.8.25. *The state vector at time t is given by*

$$x_t = \begin{pmatrix} s_t \\ \vdots \\ s_{t+n-1} \end{pmatrix} \in \mathbb{F}^n,$$

and the transition matrix by

$$A = \begin{pmatrix} 0 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 0 & 1 \\ -b_0 & \cdots & \cdots & -b_{n-1} \end{pmatrix} \in \mathbb{F}^{n \times n}.$$

Then $(*)$ has a first order description

$$x_{i+1} = Ax_i, \qquad i \in \mathbb{N}.$$

Thus $x_1, \ldots, x_{q^n-1}$ appear once and only once and pass through all non-zero vectors in $\mathbb{F}_q^n$.

Note that $A^{q^n-1} = I_n$!

But there is a great weakness of this system! Assume an attacker has access to $s_t, \ldots, s_{t+2n-1}$, which for example can be gained from a plaintext attack. From $(*)$ it follows that

$$\begin{pmatrix} s_t & s_{t+1} & \cdots & s_{t+n-1} \\ s_{t+1} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ s_{t+n-1} & \cdots & \cdots & s_{t+2n-1} \end{pmatrix} \begin{pmatrix} b_0 \\ \vdots \\ \vdots \\ b_{n-1} \end{pmatrix} = \begin{pmatrix} -s_{t+n} \\ -s_{t+n+1} \\ \vdots \\ -s_{t+2n-1} \end{pmatrix}.$$

Solving this system reveals $b_0, \ldots, b_{n-1}$ and hence also φ and f.

The *cost* of this attack: A naive Gauss elimination requires $\mathcal{O}(n^3)$ field operations. In 1969 Berlekamp and Massey came up with an algorithm to solve this linear problem in $\mathcal{O}(n^2)$ field operations! Thus, after 1969 everybody stopped using this system.

After 1969 the interest in stream ciphers remained intense. Instead of linear recurrence sequences, *nonlinear recurrence sequences* were consideres. For this let $f \in \mathbb{F}[x_1, \dots, x_n]$; then

$$s_{n+i} = f(s_{n+i-1}, \dots, s_i), \qquad i \in \mathbb{N}$$

defines a nonlinear recurrene sequence.

As an example, consider $s_{i+3} = 3s_{i+2}s_{i+1} + 6s_i^2 s_{i+2}$ over $\mathbb{F}_{19}$. In the area of nonlinear recurrence sequences, still many problems are not solved. A standard reference for such recurrence sequences is the book by Rainer Rueppel from 1986. One of the most famous nonlinear stream ciphers is probably MD5.

1.8.2 Block Ciphers

A *block cipher* is a secret key system where $\mathcal{M}$, $\mathcal{K}$ and $\mathcal{C}$ are finite sets. A famous example is DES, the data encryption standard introduced in 1975. There we have

$$|\mathcal{M}| = |\mathcal{C}| = 2^{64} \quad \text{and} \quad |\mathcal{K}| = 2^{56}.$$

Remark 1.8.26. *Under the old ASCII code,* $2^7 = 128$ *type writer symbols are encoded in*

$$\{x \in \mathbb{Z}_2^8 \mid \sum x_i = 0\}.$$

For example,

$$A \mapsto 01000001, \qquad a \mapsto 11100001, \qquad 0 \mapsto 00110000.$$

Thus eight ASCII symbols correspond to an element in $\mathbb{Z}_2^{56}$ *when the check digit is omittet. Because of advancements in computers, DES became obsolte in the mid 90'th.*

On November 26, 2001 the National Institute for Standards and Technology (NIST) adopted the Rijndael system as the advanced encryption standard AES. The inventors of the systems are Vincent Rijmen and Joan Daemen from Belgium.

We want to sketch the system. Consider the polynomial ring $\mathbb{Z}_2[x, y, z]$. Let $\mu := z^8 + z^4 + z^3 + z + 1$ (an irreducible element of $\mathbb{Z}_2[z]$) and $I = \langle \mu, x^4 + 1, y^4 + 1 \rangle$. In Rijndael,

$$\mathcal{M} = \mathcal{C} = \mathcal{K} = R := \mathbb{Z}_2[x, y, z]/I.$$

Note that R has a $\mathbb{Z}_2$-basis given by $\{x^i y^j z^k \mid 0 \le k < 8,\ 0 \le i, j < 4\}$ and thus $|R| = 2^{128}$. Moreover $\mathbb{F} = \mathbb{Z}_2[x]/\langle \mu \rangle$ is a field of $2^8 = 256$ elements.

Define a permutation of $\mathbb{F}$ through $\varphi : \mathbb{F} \to \mathbb{F}$, where $\varphi = \varphi_3 \circ L \circ \varphi_1$ and

$$\begin{aligned}
\varphi_1 : \mathbb{F} \to \mathbb{F}, &\quad f \mapsto \begin{cases} f^{-1} & \text{if } f \neq 0, \\ 0 & \text{if } f = 0, \end{cases} \\
L : \mathbb{F} \to \mathbb{F}, &\quad f \mapsto (z^4 + z^3 + z^2 + z + 1) \cdot f \mod z^8 + 1, \\
\varphi_3 : \mathbb{F} \to \mathbb{F}, &\quad f \mapsto (z^6 + z^5 + z + 1) + f.
\end{aligned}$$

In practice φ is stored via a lookup table. We want to use the following notation for an element $r \in R = \mathbb{Z}_2[x, y, z]/I$ in the following:

$$r = \sum_{i=0}^{3} \sum_{j=0}^{3} r_{ij} x^i y^j = \sum_{i=0}^{3} r_j y^j, \qquad \text{where } r_{ij} \in \mathbb{F}_{256} \text{ and } r_j \in \mathbb{F}_{256}[x]/\left\langle x^4 + 1 \right\rangle.$$

The encryption algorithm works like this: Alice and Bob exchange a secret key $k \in R$. In a first step, they both do a secret key expansion, computing 11 secret keys $k^{(t)}$, $t = 0, \dots, 10$, by

$$k^{(0)} = k, \quad k_0^{(t+1)} = \sum_{i=0}^{3} \varphi(k_{i,3}^{(t)}) x^{i+3} + z^t + k_0^{(t)}, \quad k_i^{(t+1)} = k_{i-1}^{(t+1)} + k_i^{(t)}$$

for $t = 0, \ldots, 9$ and $i = 1, 2, 3$. If now Alice wants to send the message $m \in R$, she computes

$$\begin{aligned} m^{(0)} &:= m + k^{(0)}, \\ m^{(t+1)} &:= \gamma \sum_{i=0}^{3} \sum_{j=0}^{3} \varphi(m_{ij}^{(t)}) x^i y^{3i+j} + k^{(t+1)} \quad \text{for } t = 0, \ldots, 8, \\ c := m^{(10)} &:= \sum_{i=0}^{3} \sum_{j=0}^{3} \varphi(m_{ij}^{(9)}) x^i y^{3i+j} + k^{10}, \end{aligned}$$

where $\gamma = (z+1)x^3 + x^2 + x + z \in R$.

1.9 Public Key Systems Based on the Discrete Logarithm Problem in a Finite Group

Let G be a finite group, $\alpha \in G$ an element of *finite order* $\operatorname{ord}(\alpha) = n$. Consider the cyclic subgroup $\langle\alpha\rangle = \{e, \alpha, \alpha^2, \ldots, \alpha^{n-1}\}$.

Definition 1.9.1. *For $\beta \in G$ one defines the* discrete logarithm of β with base α *as a number $a \in \mathbb{Z}$ (if exists) such that $\alpha^a = \beta$. In this case one writes $a = \log_\alpha \beta$.*

Remarks 1.9.2.

(1) *The discrete logarithm $\log_\alpha \beta$ exists if and only if $\beta \in \langle\alpha\rangle$.*

(2) *If $\log_\alpha \beta$ exists then it is "multi-valued"; this means that if $a = \log_\alpha \beta$, then also $a + kn$, $k \in \mathbb{Z}$ are discrete logarithms of β with base α.*

To be more exact, the set $\{a + kn\}$ gives all *discrete logarithms of β with base α.*

(3) *If $\beta \in \langle\alpha\rangle$ define the unique integer $a \in \{0, 1, \ldots, n-1\}$ satisfying $\alpha^a = \beta$ as the* principal value *of $\log_\alpha \beta$.*

Lemma 1.9.3 (Calculation rules).

(1) *It is $\log_\alpha \beta^k \equiv k \log_\alpha \beta \pmod{n}$ for every $k \in \mathbb{Z}$ and $\beta \in \langle\alpha\rangle$.*

(2) *It is $\log_\alpha(\beta_1\beta_2) \equiv \log_\alpha \beta_1 + \log_\alpha \beta_2 \pmod{n}$ for every $\beta_1, \beta_2 \in \langle\alpha\rangle$.*

In 1976 Diffie and Hellmann proposed a secret key exchange which was based on the hardness of the *discrete logarithm problem* (DLP):

1) Alice and Bob agree on a group G and $g \in G$.

2) Alice picks $a \in \mathbb{N}$ and computes g^a.

3) Bob picks $b \in \mathbb{N}$ and computes g^b.

4) Alice and Bob exchange g^a and g^b.

5) Both Alice and Bob can compute $g^{ab} = (g^a)^b = (g^b)^a$.

Note that to compute g^{ab} when one only knows G, g, g^a and g^b, one has to solve a discrete logarithm problem to either find a or b.

Remark 1.9.4. *Alice and Bob pick $a, b \geq 2^{100}$ so that Eve (the eavesdropper) cannot simply find a or b by exhaustive search. (Using consecutive squaring one can compute g^a and g^b easily, using $\mathcal{O}(\log a)$ group operations.)*

The major *drawback* is that this system does not gives a one-way trapdoor function!

In 1985, El Gamal showed how to create a one-way trapdoor function from difficult discrete logarithm problems:

- Alice choses $\alpha \in G$ and $\beta = \alpha^a$ for some $a \in \mathbb{N}$.
- She makes (G, α, β) public, and keeps a secret.
- Encryption is done via the randomized "function" $\varphi : G \to G \times G$, $m \mapsto (\alpha^k, m\beta^k)$ where $k \in \mathbb{Z}$ is randomly chosen by Bob.
- Decryption is done via $\psi : G \times G \to G$, $(c_1, c_2) \mapsto c_2 c_1^{-a}$.
  If $c_1 = \alpha^k$ and $c_2 = m\beta^k$, we have $c_2 c_1^{-a} = m\beta^k\alpha^{-ak} = m\alpha^{ak-ak} = m$.

Note that Alice cannot recover k without solving a discrete logarithm problem herself.
This leads to the

Question 1.9.5. *How hard is DLP?*

A "*pure math answer*": Consider the homomorphism $\psi : \mathbb{Z} \to G$, $a \mapsto \alpha^a$ and let $n = \operatorname{ord} \alpha$. By the homomorphism theorem, $\mathbb{Z}_n = \mathbb{Z}/\ker\psi \cong \langle a \rangle$ and thus there exists an isomorphism $\rho : \langle a \rangle \to \mathbb{Z}_n$, $\alpha^a \mapsto a + n\mathbb{Z}$. If Eve wants to compute $\log_\alpha \beta = a$, she can do this by applying ρ to β, since $\log_\alpha \beta + n\mathbb{Z} = \rho(\beta)$.

But of course this does not really works, since to be able to *compute* the map ρ one has to solve a discrete logarithm problem every time!

Note that in $\mathbb{Z}$, solving $ax \equiv b \pmod{n}$ is easy. Thus the discrete logarithm problem in the additive group of $\mathbb{Z}_n$ is easy.

There are many groups where the discrete logarithm problem has been studied in literature. We want that G has a cyclic subgroup of order at least 2^{100}.

Examples 1.9.6.

(1) *$\mathbb{Z}_n^*$; this is cyclic of order $\phi(n)$;*

(2) *$\mathbb{F}_{p^n}$; this is cyclic of order $p^n - 1$;*

(3) *$GL_n(\mathbb{F}_q)$, the invertible $n \times n$-matrices over $\mathbb{F}_q$;*

(4) *$E(\mathbb{F}_q)$, the $\mathbb{F}_q$-rational points of an elliptic curve;*

(5) *Jacobians of varieties.*

1.9.1 Solving the Discrete Logarithm Problem

We first want to concentrate on methods to *solve* the discrete logarithm problems in (fairly) arbitrary groups.

(1) *Exhaustive search*:

For $i = 1, 2, 3, \ldots$ compute $\alpha^i \stackrel{?}{=} \beta$. If $n = |\langle \alpha \rangle|$ this has a cost of $\mathcal{O}(n)$ group operations.

(2) *Baby-step Giant-step* method:

This method was invented by Shanks.

The *baby step*: For some number m produce a look-up table and store it in the computer:

$$\{(i, \alpha^i) \mid 0 \le i \le m\}.$$

The *giant step*: Compute $\beta(\alpha^{-m})^j$ for $j = 1, 2, 3, \ldots$ and compare the result with the look-up table. If $\beta(\alpha^{-m})^j = \alpha^i$ for $0 \le i \le m$¡ then $\beta = \alpha^{i+mj}$ and thus we are done.

As an *example*, take $m = \lfloor \sqrt{n} \rfloor$, i. e. the largest integer smaller or equal to $\sqrt{n}$. Then $\mathcal{O}(\sqrt{n})$ numbers have to be stored, and $\mathcal{O}(\sqrt{n})$ group multiplications have to be performed.

(3) *Pohlig-Hellmann algorithm*:

Assume G has order $n = p_1^{s_1} \cdots p_r^{s_r}$, where the p_i are pairwise distinct primes, and $p_1, \ldots, p_r \le B$ for a fairly small bound B. (Thus n is B-smooth.) Under this condition the discrete logarithm can be computed iteratively:

Assume $\alpha^x = \beta$. For $i = 1, \ldots, r$ let $x_i = x \mod p_i^{s_i}$. If the x_i are known, by the Chinese Remainder Theorem also x is known. Now fix one i.

Let $x_i = \sum_{j=0}^{s_i - 1} \ell_j p_i^j$ with $0 \le \ell_j > p_i$. Pohlig-Hellmann computes iteratively $\ell_0, \ell_1, \ldots, \ell_{s_i - 1}$ by the following method:

Establish a look-up table for the p-th root of unity of α,

$$\{(k, \alpha^{k \cdot \frac{n}{p_i}}) \mid k = 0, \ldots, p_i - 1\}.$$

In order to find ℓ_0, compute $\beta^{\frac{n}{p_i}}$. Then $\beta^{\frac{n}{p_i}} = \alpha^{x \cdot \frac{n}{p_i}}$, and $x\frac{n}{p_i} \equiv k\frac{n}{p_i} \pmod{n}$ if and only if $x \equiv k \pmod{p_i}$. Since $x \equiv x_i \equiv \ell_0 \pmod{p_i}$ we can get ℓ_0 from the look-up table.

Now $\beta\alpha^{-\ell_0} = \alpha^{x - \ell_0}$, so we also get $(\beta\alpha^{-\ell_0})^{\frac{n}{p_i^2}} = (\alpha^{x-\ell_0})^{\frac{n}{p_i^2}} = (\alpha^{p\ell_1})^{\frac{n}{p_i^2}}$, since $x - \ell_0 \equiv p_i\ell_1 \pmod{\frac{n}{p_i^2}}$. Thus we can get ℓ_1 from the look-up table. Continuing in this fashion delivers $\ell_2, \ell_3, \ldots, \ell_{s_i - 1}$.

This practically only works if the look-up talbe has reasonable size, and thus $B \leq 2^{30}$ (for example). In order to avoid Pohlig-Hellmann the group order should be divisible by a prime $p \geq 2^{50}$. In case $G = \mathbb{Z}_p^*$ pick primes p where $\frac{p-1}{2}$ is prime as well (these are called safe primes).

Remark 1.9.7. *The running time of Pohlig-Hellmann is* $\mathcal{O}(\sum_{i=1}^{r} s_i(\log n + \sqrt{p_i}))$.

(4) *Index calculus*:

Assume the group G has some *factor base* $S = \{p_1, \ldots, p_t\}$. The elements $p_1, \ldots, p_t$ are group elements such that for an arbitrary chosen $g \in G$ there is a good chance to write $g = p_1^{d_1} \cdots p_t^{d_t}$, where the $d_i \in \mathbb{Z}$. (A good chance is here for example a chance of at least 0.01%). In such a situation search for some k such that

$$\alpha^k = p_1^{d_1} \cdots p_t^{d_t}.$$

Thus we have

$$k \equiv \sum_{i=1}^{t} d_i \log_\alpha p_i \pmod{\operatorname{ord}(\alpha)}$$

Assume that $m \geq t$ numbers k_i could be found such that

$$k_i \equiv \sum_{j=1}^{t} d_{ij} \log_\alpha p_j \pmod{\operatorname{ord}(\alpha)}, \qquad i = 1, \ldots, m,$$

that is

$$\alpha^{k_i} = p_1^{d_{i1}} \cdots p_t^{d_{it}}, \qquad i = 1, \ldots, m.$$

This gives a linear system of equations

$$\begin{pmatrix} k_1 \\ \vdots \\ k_m \end{pmatrix} = \begin{pmatrix} d_{11} & \cdots & d_{t1} \\ \vdots & \ddots & \vdots \\ d_{1m} & \cdots & d_{tm} \end{pmatrix} \begin{pmatrix} \log_\alpha p_1 \\ \vdots \\ \log_\alpha p_t \end{pmatrix} \pmod{\operatorname{ord}(\alpha)},$$

which can (if the rank is high enough) be used to compute

$$\log_\alpha p_1, \ \ldots, \ \log_\alpha p_t \pmod{\operatorname{ord}(\alpha)}.$$

In order to compute $\log_\alpha \beta$, search for an $\ell \in \mathbb{Z}$ such that

$$\alpha^{\ell + \log_\alpha \beta} = \beta \alpha^\ell = p_1^{e_1} \cdots p_t^{e_t}, \qquad e_i \in \mathbb{Z}.$$

Then

$$\log_\alpha \beta \equiv -\ell + \sum_{i=1}^{t} e_i \log_\alpha p_i \pmod{\operatorname{ord}(\alpha)},$$

and we are done!

Inside $\mathbb{Z}_p^*$, one could take the first t primes as a factor base, like $S = \{2, 3, 5, 7, 11, 13, \ldots\}$.

Example 1.9.8. *Let* $G = \mathbb{Z}_{220}^*$. *Take* $S = \{2, 3, 5, 7, 11\}$, $\alpha = 6$ *and* $\beta = 13$. *Find some* $a \in \mathbb{Z}$ *such that* $\alpha^a \equiv \beta \pmod{229}$.

We have $6^1 = 2 \cdot 3$, $6^{12} \equiv 165 = 3 \cdot 5 \cdot 11 \pmod{229}$, *which gives the relations*

$$1 \equiv \log_\alpha 2 + \log_\alpha 3 \pmod{228}$$
$$\textit{and} \qquad 12 \equiv \log_\alpha 3 + \log_\alpha 5 + \log_\alpha 11 \pmod{228}.$$

Moreover we get

$$6^{18} \equiv 176 = 2^4 \cdot 11 \pmod{229},$$
$$6^7 \equiv 98 = 2 \cdot 7^2 \pmod{229}$$
$$\textit{and} \qquad 6^x \equiv \ldots \pmod{229},$$

thus giving the relations

$$\begin{aligned} 18 &\equiv 4 \cdot \log_\alpha 2 + \log_\alpha 11 \pmod{228}, \\ 7 &\equiv \log_\alpha 2 + 2 \log_\alpha 7 \pmod{228} \\ \textit{and} \qquad \cdots &\equiv \ldots \pmod{228}. \end{aligned}$$

From this we get

$$\begin{aligned} &\log_\alpha 2 = 21, && \log_\alpha 3 = 208, && \log_\alpha 5 = 98, \\ &\log_\alpha 7 = 107 && \textit{and} && \log_\alpha 11 = 162. \end{aligned}$$

Now search for $\ell \in \mathbb{Z}$ *such that* $13 \cdot 6^\ell \mod 229$ *factors over* S*. We get*

$$13 \cdot 6^2 \equiv 2 \cdot 5 \pmod{229},$$

and thus

$$\log_\alpha 13 = \log_\alpha 2 + \log_\alpha 5 - \log_\alpha 6^2 \mod 228 = 21 + 98 - 2 \mod 228 = 117.$$

The following example illustrating how index calculus can be done over $\mathbb{F}_q$ is taken from the [MvOV96].

Example 1.9.9 (Index calculus over $\mathbb{F}_q$). *Let* $G = \mathbb{F}_{128}^* = (\mathbb{Z}_2[Z]/(f))^*$*, where* $f = Z^7 + Z + 1 \in \mathbb{Z}_2[Z]$*. As a factor base take al irreducible polynomials of degree at most 3, i. e.*

$$S = \{\underbrace{z}_{=:p_1}, \underbrace{z+1}_{=:p_2}, \underbrace{z^2+z+1}_{=:p_3}, \underbrace{z^3+z+1}_{=:p_4}, \underbrace{z^3+z^2+1}_{=:p_5}\}.$$

Denote the image of Z *in* $\mathbb{Z}_2[Z]/(f)$ *by* z*. Take* $\alpha = z$*; since* $|\mathbb{F}_{128}^*| = 127$ *is prime, any* $\alpha \neq 1$ *is a generator of* G*. Let further be* $\beta = z^4 + z^3 + z^2 + z + 1$*; the task is to find* $\log_\alpha \beta$*.*

Let $\ell_i = \log_\alpha p_i$*. To faktor an* $x \in \mathbb{F}_{128}^*$*, we treat it and the* p_i*'s as polynomials in* $\mathbb{Z}_2[Z]$ *and try to factor there. We get*

$$\begin{aligned} \alpha^1 &= z = p_1, \\ \alpha^{18} &= z^6 + z^4 = p_1^4 p_2^2, \\ \alpha^{45} &= z^5 + z^2 + z1 = p_2^2 p_4, \\ \alpha^{72} &= z^6 + z^5 + z^2 = p_1^2 p_2^2 p_3, \\ \alpha^{105} &= z^6 + z^5 + z^4 + z = p_1 p_2^2 p_5, \\ \alpha^{121} &= z^6 + z^5 + z^3 + z^2 + z + 1 = p_4 p_5. \end{aligned}$$

We get the linear system

$$\begin{pmatrix} 1 \\ 18 \\ 45 \\ 72 \\ 105 \\ 121 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 4 & 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 0 \\ 2 & 2 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} \ell_1 \\ \vdots \\ \ell_5 \end{pmatrix},$$

resulting in

$$\ell_1 = 1,\ \ell_2 = 7,\ \ell_3 = 56,\ \ell_4 = 31 \textit{ and } \ell_5 = 90.$$

Now we find

$$\beta\alpha^{66} = z^5 + z^3 + z = p_1 p_3^2,$$

and thus

$$\log_\alpha \beta = \ell_1 + 2\ell_3 - 66 \mod 127 = 47.$$

Example 1.9.10 (A case study). *The* Digital Signature Standard *(1991) is based on a discrete logarithm problem in $\mathbb{F}_p^*$, where*

$$2^{511+64t} \leq p \leq 2^{512+64t}, \qquad t = 0, \ldots, 8;$$

here t is a security parameter. The underlying algorithm of the standard is known as the Digital Signature Algorithm (DSA).

In 1991 index calculus was already known, and we want to show what effect index calculus had when the possible sizes of p were specified. Assume that $p \approx 2^{1000}$ (that roughly corresponds to security parameter between 7 and 8). To do index calculus, we need a factor base $S = \{2, 3, 5, \ldots, p_t\}$, where $p_t \leq B$ for some smoothness bound B.

A natural question in this case is: what is the probability that a random number $x \in \{1, 2, \ldots, p-1\}$ is B-smooth? This is answered by the following theorem:

Theorem 1.9.11 (Norton (1971), Canfield, Erdős, Pomerance (1983)). *Let N and r be positive reals satisfying*

$$B := N^{1/r} \geq \log N.$$

Then the number of $x \in \mathbb{N}$, $x \leq N$ which are B-smooth is given by

$$N \cdot r^{-r+o(r)}, \qquad \text{where} \quad \lim_{N \to \infty} \frac{o(r)}{r} = 0.$$

Example 1.9.12 (Continuing Example 1.9.10).

Let $p \approx 2^{1000}$ and $r = 20$. Then $B \approx 2^{1000/20} = 2^{50}$, and it is expected that one out of $20^{20} \approx 10^{26}$ numbers can be factored over the base $\{p_1, \ldots, p_t\}$, where $p_t \leq 2^{50}$. By the Prime Number Theorem, $t \approx \frac{2^{50}}{\log 2^{50}} \approx 2^{45}$. Thus to use index calculus here, one has to compute and store a $2^{45} \times 2^{45}$ matrix and solve the associated linear system, which is even out of range for future computers.

Even if the lowest security parameter is chosen, index calculus is no real threat to the security. But for numbers $\approx 2^{100}$, index calculus is well suited. Currently, the *Generalized Number Field Sieve* is the best algorithm for solving the DLP in this case.

(5) In the case there is no factor base known, the best known method to solve a DLP is the *Pollard ρ method*:

For this let $G = \langle \alpha \rangle$ be a cyclic group and $\beta \in G$. As usual, $k := \log_\alpha \beta$ is wanted. For simplicity let $|G| = p$ be prime. We search for exponents (x_i, y_i) and (x_j, y_j) such that

$$\alpha^{x_i} \beta^{y_i} = \alpha^{x_j} \beta^{y_j}.$$

If such a relation is found, we have

$$\alpha^{x_i - x_j} = \beta^{y_j - y_i} = \alpha^{k(y_j - y_i)} \qquad \text{and thus} \qquad x_i - x_j \equiv k(y_j - y_i) \pmod{p}.$$

Thus if $y_j - y_i$ is invertible in $\mathbb{Z}_p$, we can compute

$$k = (x_i - x_j)(y_j - y_i)^{-1} \mod p.$$

Question 1.9.13. *How many exponents (x_i, y_i), $i = 1, 2, 3, \ldots$ should be randomly checked until a collision becomes likely?*

This problem is known as the *birthday problem*. An *approximate* answer is: given $\sqrt{p}$ randomly chosen (x_i, y_i) will provide a collision with probability at least $\frac{1}{2}$.

In this way $\sqrt{p}$ elements of the form $\alpha^{x_i} \beta^{y_i}$ have to be computed and stored! (This is very similar to the baby-step giant-step algorithm.)

Pollard showed how to eliminate the storage problem: Define a recurrence sequence

$$x_{i+1} = f(x_i, y_i),\ y_{i+1} = g(x_i, y_i),\ (x_0, y_0) = (1, 1)$$

such that most indices $(x, y) \in \mathbb{Z}_p^2$ are visited in the sequence $(x_i, y_i)_{i\in\mathbb{N}}$. Consider the sequence of group elements

$$\{g_i := \alpha^{x_i}\beta^{y_i} \mid i \in \mathbb{N}\}.$$

Since $(g_i)_i$ must be ultimately periodic, assume it has pre-period m and period n. In order to find a collision, not all elements have to be stored, instead just store the current element of the sequence

$$(g_i, g_{2i})_i, \quad i \in \mathbb{N}.$$

For one i there will be $g_i = g_{2i}$, and thus one can find a collision this way! The costs are $\mathcal{O}(\sqrt{p})$ time consumption, and (almost) no storage consumption.

(The name "Pollard ρ" originates to that if one wants to illustrate the algorithm graphically, one draws a circle with a line attached, which if correctly orientated looks like the Greek letter ρ.)

1.10 An Introduction to Elliptic Curves

1.10.1 Affine Curves

Let $p = \sum a_{ij}x^i y^i \in \mathbb{F}[x, y]$ be a polynomial in two variables. Then

$$C := V(p) := \{(x, y) \in \mathbb{F}^2 \mid p(x, y) = 0\}$$

is called an *affine curve* of *degree* $\deg C := \max\{i + j \mid a_{ij} \neq 0\}$. (The "$V$" stands for *variety*.)

Example 1.10.1. *The curve given by* $\frac{x^2}{a} + \frac{y^2}{b} - 1 = 0$ *is an ellipse, and has degree* 2.

Curves of degree 1, 2, 3, 4 and 5 are called *lines*, *conics*, *cubics*, *quartics* and *quintics*, respectively. One says a curve $p(x, y) = 0$ is *irreducible* if p as a polynomial is irreducible. A point (α, β) on the curve is called *smooth* if

$$\left(\frac{\partial p}{\partial x}(\alpha, \beta), \frac{\partial p}{\partial y}(\alpha, \beta)\right) \neq (0, 0);$$

otherwise one says that (α, β) is *singular*. A curve C is called *smooth* if all points on it are smooth.

1.10.2 Bezóut's Theorem for Curves

Theorem 1.10.2 (Bezóut). *Let* C_1 *and* C_2 *be irreducible curves of degree* d_1 *and* d_2*, and assume that* $C_1 \neq C_2$*. Then*

$$|C_1 \cap C_2| \leq d_1 d_2.$$

Over the algebraic closure of $\mathbb{F}$*, when computed with multiplicities*[8] *and points at infinity (we will see later what these are) we have that the number of common points of* C_1 *and* C_2 *is exactly* $d_1 d_2$*.*

1.10.3 Projective Plane

Definition 1.10.3. *For a field* $\mathbb{F}$ *let*

$$\mathbb{P}^2 := \mathbb{P}^2_{\mathbb{F}} := \{(\alpha, \beta, \gamma) \in \mathbb{F}^3 \setminus \{(0, 0, 0)\}\}/_\sim,$$

where $\sim$ *is an equivalence relation defined by*

$$(\alpha, \beta, \gamma) \sim (\alpha', \beta', \gamma') \iff \exists \lambda \in \mathbb{F}^* : (\lambda\alpha, \lambda\beta, \lambda\gamma) = (\alpha', \beta', \gamma').$$

Then $\mathbb{P}^2_{\mathbb{F}}$ *is called the* projective plane *over* $\mathbb{F}$*. A point* $[(\alpha, \beta, \gamma)]_\sim$ *will be written simply as* (α, β, γ)*. If* $\gamma \neq 0$*,* (α, β, γ) *is called* finite*, otherwise it is called* infinite *or* point at infinity*.*

One has a one-to-one correspondence between $\mathbb{F}^2$ and the finite points $\mathbb{P}^2$ given by

$$\varphi : \mathbb{F}^2 \to \mathbb{P}^2, \quad (x, y) \mapsto (x, y, 1).$$

This can be viewed graphically as in picture 1.1. The infinite points have the form $(\alpha, \beta, 0)$, and can be thought as lines in the x-y-plane in the above picture. The projective plane can be thought as some kind of "closure" of the affine plane $\mathbb{F}^2$, and it has neat properties which the affine plane does not have; for example, every two distinct lines share exactly one point, and every two distinct points share exactly one line.

Let $p \in \mathbb{F}[x, y]$ be a polynomial of degree d describing the affine curve $V(p)$.

Definition 1.10.4. *For* $p \in \mathbb{F}[x, y]$ *of degree* d *denote the polynomial* $\hat{p} \in \mathbb{F}[x, y, z]$ *defined by* $\hat{p}(x, y, z) := z^d p(\frac{x}{z}, \frac{y}{z})$ *as the* homogenization *of* p*.*

Note 1.10.5.

[8] Defining the multiplicity for an intersection point of two curves is not trivial, and at the moment we just want to illustrate it with an example: if a line is a tangent to a curve, the intersection has multiplicity two, whereas the multiplicity is one if the line intersects the curve with another angle.

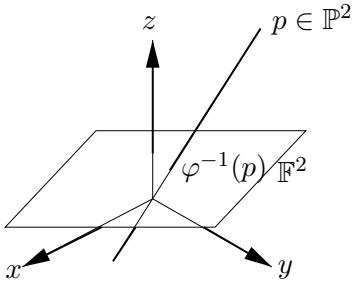


Figure 1.1: Mapping a finite point of $\mathbb{P}^2$ to $\mathbb{F}^2$

(1) *We have* $\hat{p}(\lambda x, \lambda y, \lambda z) = \lambda^d \hat{p}(x, y, z)$ *for all* $\alpha, x, y, z \in \mathbb{F}$.

(2) *By substituting* $z \mapsto 1$, $\hat{p}$ *reduces to* p*: it is* $\hat{p}(x, y, 1) = p(x, y)$ *for all* $x, y \in \mathbb{F}$.

Examples 1.10.6.

(1) *The parabola* $y = x^2$ *is defined by* $p := y - x^2 \in \mathbb{F}[x, y]$. *For it we get* $\hat{p} = yz - x^2$.

(2) *If* $p = 3y^2 + x^3 + xy + 5$, *we have* $\hat{p} = 3y^2z + x^3 + xyz + 5z^3$.

The homogenous form $\hat{p}$ defines the projective curve

$$C := V(\hat{p}) := \{(\alpha, \beta, \gamma) \in \mathbb{P}^2 \mid \hat{p}(\alpha, \beta, \gamma) = 0\}.$$

(Note that $\hat{p}(\alpha, \beta, \gamma) = 0$ is well defined for any $(\alpha, \beta, \gamma) \in \mathbb{P}^2$ by the second note: for $\lambda \neq 0$ we have $\hat{p}(\alpha, \beta, \gamma) = 0$ if and only if $\hat{p}(\lambda\alpha, \lambda\beta, \lambda\gamma)$.) Now it is

$$V(p) \cup \{(\alpha, \beta, 0) \in \mathbb{P}^2 \mid \hat{p}(\alpha, \beta, 0)\} = V(\hat{p}),$$

thus the finite points of $V(\hat{p})$ are exactly all the points of $V(p)$!

Example 1.10.7. *Let* $p = y - x^2$, $\hat{p} = yz - x^2$. *Then* $V(\hat{p}) = V(p) \cup \{(0, 1, 0)\}$. *Take a look at figure 1.2.*

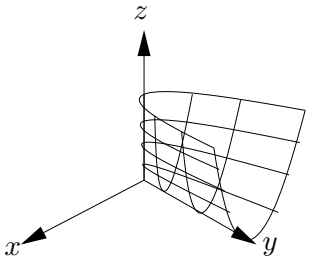


Figure 1.2: The parabola defined by $yz - x^2 = 0$

We can now restate Bezóut's theorem in its full generality:

Theorem 1.10.8 (Bezóut). *Assume* $\hat{p}_1, \hat{p}_2 \in \mathbb{F}[x, y, z]$ *are homogenous forms of degree* d_1, d_2 *defining distinct curves* $C_i = V(\hat{p}_i)$. *Then, when counted with multiplicity,* C_1 *and* C_2 *intersect in exactly* $d_1 d_2$ *points over the algebraic closure of* $\mathbb{F}$.

Example 1.10.9.

(1) *Let* $p = y - x^2$ *a parabola and* $q = x$ *the* x*-axis. Then* $\hat{p} = yz - x^2$ *and* $\hat{q} = x$. *The system*

$$yz - x^2 = 0, \qquad x = 0$$

has two solutions in $\mathbb{P}^2$ *given by*

$$(0, 0, 1) \qquad \textit{and} \qquad (0, 1, 0).$$

(2) *Let $p = x+2y$ and $q = x+2y+1$ be two parallel lines. We have $\hat{p} = x+2y$ and $\hat{q} = x+2y+z$, and we get the solution*

$$(2, -1, 0),$$

which is a point at infinity. It can be geometrically interpreted as pointing in the direction of the lines; in fact, all points at infinity can be interpreted as directions, or as points infinitely far away from the origin into a direction.

We want to give an outline of the proof of the Theorem of Bezóut. For this, we need a technical lemma:

Lemma 1.10.10. *Let $f = \sum_{i=0}^{n} f_i x^i$ and $g = \sum_{i=0}^{m} g_i x^i$ be two polynomials of degree n and m, respectively, where $f, g \in \mathbb{F}[x]$. Define*

$$S(f,g) := S := \begin{pmatrix} f_n & 0 & \cdots & 0 & g_m & & 0 \\ \vdots & f_n & \ddots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 & \vdots & & g_m \\ f_0 & \vdots & & f_n & \vdots & & \vdots \\ 0 & f_0 & & \vdots & g_0 & & \vdots \\ \vdots & \ddots & \ddots & \vdots & & \ddots & \vdots \\ 0 & \cdots & 0 & f_0 & 0 & & g_0 \end{pmatrix} \in \mathbb{F}^{(n+m)\times(n+m)}.$$

The determinant of S is called the resultant $\operatorname{Res}(f,g)$. *Then $\operatorname{Res}(f,g) = \det S = 0$ if and only if $\gcd(f,g) \neq 1$.* resultant

Proof sketch for the lemma. The matrix S defines a map $\mathbb{F}^m \times \mathbb{F}^n \to \mathbb{F}^{m\times n}$, $(a,b) \mapsto af + bg$, where $\mathbb{F}^k$ is interpreted as the vector space of polynomials over $\mathbb{F}$ of degree strict less than k. The proof can be done as follows:

If f and g are coprime, one needs to show that there is a Bezóut equation $af + bg = 1$ for f and g where $\deg a < m$ and $\deg b < n$. Thus $1 = (0, \ldots, 0, 1)$ is in the image of S. Conversely, if f and g are not coprime, then there are no $a, b \in \mathbb{F}[x]$ such that $1 = af + bg$, and thus 1 is not in the image of S. We can conclude by showing that $\det S$ is zero if and only if $1 = (0, \ldots, 0, 1)$ is not in the image of S. □

Note 1.10.11. *Note that the lemma can also be applied if $\mathbb{F}$ is not a field but an integer domain, and if the polynomials are monic: by the Lemma of Gauss monic polynomials over $\mathbb{F}$ are irreducible iff they are prime iff they are irreducible over the field of fractions of $\mathbb{F}$ iff they are prime over the field of fractions of $\mathbb{F}$.*

Thus if two monic polynomials f and g are coprime over $\mathbb{F}$, they are also coprime over the field of fractions of $\mathbb{F}$. Since the determinant of S is the same over $\mathbb{F}$ and its field of fractions, we can show this claim by switching from $\mathbb{F}$ to its field of fractions.

Proof sketch of the Theorem of Bezóut. Write $\hat{p}_i = \sum_{j=0}^{d_i} p_{i,j} x^j$, where $p_{i,j} \in \mathbb{F}[y,z]$. Consider $S(\hat{p}_1, \hat{p}_2)$; this is a homogenous polynomial in $\mathbb{F}[y,z]$ of degree $n \cdot m$. By the fundamental theorem of algebra, one can factor $\det S$ (since $\det S = z^{n\cdot m}(\det S)(\frac{y}{z}, 1)$), so we get

$$\det S = \prod_{k=1}^{nm} (\beta_i z - \gamma_i y).$$

Thus we have $n \cdot m$ solutions for (y,z) such that there is a solution α_i where $(\alpha_i, \beta_i, \gamma_i) \in V(\hat{p}_1) \cap V(\hat{p}_2)$. □

1.10.4 Elliptic Curves

Definition 1.10.12. *A nonsingular projective curve having the homogenous form*

$$y^2 z + a_1 xyz + a_3 yz^2 = x^3 + a_2 x^2 z + a_4 xz^2 + a_6 z^3$$

respectively the inhomogenous form

$$y^2 + a_1 xy + a_3 y = x^3 + a_2 x^2 + a_4 x + a_6,$$

where $a_i \in \mathbb{F}$, is called an elliptic curve *in* Weierstrass form.

This curve has exactly one point at infinity, which is $\mathcal{O} := (0, 1, 0)$.

Note 1.10.13. *Since $GL_3(\mathbb{F})$ operates on $\mathbb{P}^2$, we can change bases of the elliptic curve by applying invertible transformations onto the coordinates.*

If $\mathrm{Char}(\mathbb{F}) \neq 2$*, then the Weierstrass form can be simplified by substituting $y \mapsto \frac{1}{2}(y - a_1x - a_3)$, resulting in*

$$y^2 = x^3 + \tilde{a}_2x^2 + \tilde{a}_4x + \tilde{a}_6.$$

If $\mathrm{Char}(\mathbb{F}) \neq 3$*, by substituting $x \mapsto x - \frac{a_2}{3}$ this can be more simplified to*

$$y^2 = x^3 + ax + b, \qquad a, b \in \mathbb{F}.$$

Remarks 1.10.14. *Some historical remarks: an integral of the form*

$$\int_\gamma \frac{f(z)}{\sqrt{(z - \alpha_1)(z - \alpha_2)(z - \alpha_3)}}\, dz = \int_\gamma \frac{f(z)}{y}\, dz$$

is called an elliptic integral*. The denominator (squared) is of the form*

$$y^2 = (z - \alpha_1)(z - \alpha_2)(z - \alpha_3),$$

an elliptic curve!

Question 1.10.15. *Given an equation of the form $y^2 = x^3 + ax + b$, where $a, b \in \mathbb{F}$, when does this defines an elliptic curve? I. e., when is it non-singular?*

Lemma 1.10.16. *The equation $y^2 = x^3 + ax + b$, $a, b \in \mathbb{F}$ defines an elliptic curve if and only if the* discriminant $\Delta := 4a^3 + 27b^2 \neq 0$.

Note 1.10.17. *For a homogenous curve $\hat{f} = 0$, (α, β, γ) is singular if and only if*

$$\frac{\partial \hat{f}}{\partial x}(\alpha, \beta, \gamma) = 0, \quad \frac{\partial \hat{f}}{\partial y}(\alpha, \beta, \gamma) = 0 \quad \textit{and} \quad \frac{\partial \hat{f}}{\partial z}(\alpha, \beta, \gamma) = 0.$$

Proof. We first show that the point at infinity $\mathcal{O} = (0, 1, 0)$ is smooth. We have $\hat{f} = y^2z - x^3 - axz^2 - bz^3$ and thus

$$\frac{\partial \hat{f}}{\partial x} = -3x^2 - az^2, \quad \frac{\partial \hat{f}}{\partial y} = 2zy \quad \text{and} \quad \frac{\partial \hat{f}}{\partial z} = y^2 - 2axz - 3bz^2.$$

Plugging $(0, 1, 0)$ in gives $0, 0, 1$, and thus $\mathcal{O}$ is always a smooth point. Thus we can assume $z = 1$, so we get the equations

$$3x^2 + a = 0, \quad 2y = 0 \quad \text{and} y^2 = 2ax + 3b;$$

thus $y = 0$ and we further reduce to

$$3x^2 = -a \quad \text{and} \quad (2a)x = -3b.$$

If $a = 0$, then it must be $x = 0$ and $b = 0$, and the curve $y^2 = x^3$ is singular in $(0, 0, 1)$ as one easily checks. So assume $a \neq 0$.

So the singular condition is fulfilled if and only if $\frac{27b^2}{4a^2} = 3(-\frac{3b}{2a})^2 = -a$, i. e. $27b^2 + 4a^3 = 0$. Thus if this equation is not fulfilled, the curve cannot be singular. Conversely, if $27b^2 + 4a^3 = 0$, then let $x := -\frac{3b}{2a}$; we conclude by showing that $(x, 0, 1)$ lies on the curve. But this is true, since

$$(2a)^3(x^3 + ax + b) = -27b^3 - 12a^3b + 8a^3b = -b(27b^2 + 4a^3) = 0.$$

□

1.10.5 The group law

If $P, Q \in C$ are two points on an elliptic curve C, the line through P and Q is well defined, even if $P = Q$, since C is smooth (take the tangent in that case). Define $P * Q$ as the third intersection point of C and this line; the existence is garuanteed by the Theorem of Bezóut. Even over an arbitrary field, if P and Q have coordinates in it, the third point will also have coordinates in it. Define for an P the *conjugate point* $\overline{P} := P * \mathcal{O}$. Then define

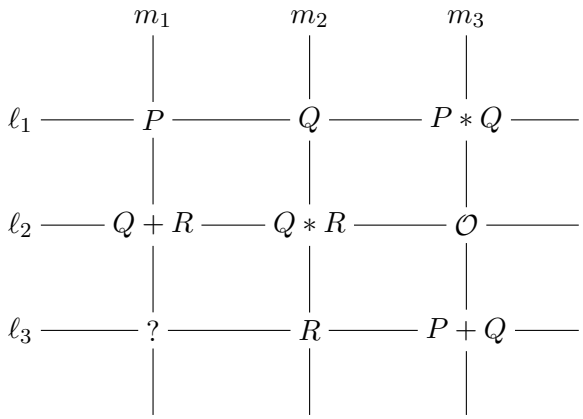
$$P + Q := \overline{P * Q}.$$

Theorem 1.10.18. *The set of points on C over an arbitrary field $\mathbb{F}$ containing the defining equation of C together with the operation $+$ has the structure of an Abelian group.*

Sketch of proof.

- The group law is surely commutative.
- The neutral element is $\mathcal{O}$, since clearly $P + \mathcal{O} = \mathcal{O} * (\mathcal{O} * P) = P$.
- The inverse is given by $-P = \overline{P}$.
- The associativity for the general case can be seen from the following (for the special cases the proof is much simplier):

  It is enough to show that $T := (P + Q) * R$ is equal to $\hat{T} := P * (Q + R)$. Take a look at the following diagram consisting of six lines ℓ_i, m_i, $i = 1, 2, 3$:



  When coming from above, one sees that the missing point is $\hat{T}$, and when coming from right, one sees that it must be T. Thus we are done.

□

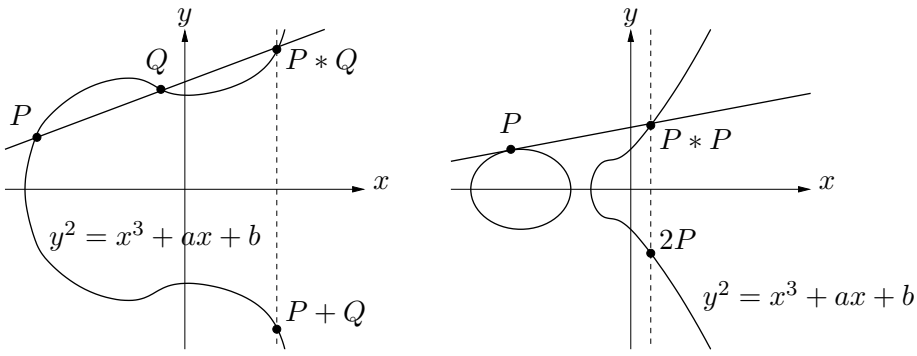


Figure 1.3: Illustration of the group law and the two typical forms of elliptic curves (Please note that these are absolutely unrealistic drawn pictures of elliptic curves. Consult your favourite plotting software to get an accurate picture :-))

Lemma 1.10.19. *Let E be an elliptic curve in $\mathbb{P}^2_{\mathbb{F}}$, given by*

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6, \qquad a_i \in \mathbb{F}.$$

Let $P_1 = (x_1, y_1), P_2 = (x_2, y_2) \in E(\mathbb{F})$. Then

$$-P_1 = (x_1, -y_1 - a_1x_1 - a_3),$$

and if $P_1 \neq -P_2$,

$$\lambda = \begin{cases} \frac{y_2 - y_1}{x_2 - x_1} & \text{if } x_1 \neq x_2, \\ \frac{3x_1^2 + 2a_2x_1 + a_4 - a_1y_1}{2y_1 + a_1x_1 + a_3} & \text{if } x_1 = x_2,\ P_1 = P_2 \end{cases}$$

and $P_1 + P_2 =: P_3 = (x_3, y_3)$, then

$$x_3 = \lambda^2 + a_1\lambda - a_2 - x_1 - x_2, \qquad y_3 = -y_1 + \lambda(x_1 - x_3).$$

Proof. The slope of the line through P_1 and P_2, respectively the slope of the tangent in $P_1 = P_2$ is given by λ. The formulas can be verified by tedious calculations which we will skip here. □

Remark 1.10.20. *If P_1 and P_2 have coordinates in a field extension of $\mathbb{F}$, then their sum $P_1 + P_2$ also has coordinates in the same extension. This motivates the following notation:*

Definition 1.10.21. *Let $E : y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6$ be an elliptic curve. If $\mathbb{F}$ is the smallest field such that all the a_i's are in $\mathbb{F}$, we say that E is* defined over $\mathbb{F}$. *We write E for $E(\overline{\mathbb{F}}) \cup \{\mathcal{O}\}$, where $\overline{\mathbb{F}}$ denotes the algebraic closure of $\mathbb{F}$. If K is a field extension of $\mathbb{F}$, we write*

$$E(K) := \{(x, y) \in K^2 \mid y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6\} \cup \{\mathcal{O}\}.$$

*The set $E(K)$ is called the set of K-*rational points.

We have seen that $(E(K), +)$ is an Abelian group for every field extension K of $\mathbb{F}$. If K is finite, $E(K)$ is obviously also finite.

We can use the group $E(\mathbb{F}_q)$ to do a Diffie-Hellmann key exchange, or to construct an ElGamal one-way trapdoor function.

This leads to the question: how difficult is the *Elliptic Curve Discrete Logarithm Problem (ECDLP)*? Namely, if $P, Q \in E(\mathbb{F}_q)$, $Q \in \langle P \rangle$, the ECDLP asks: find an $n \in \mathbb{Z}$ such that $nP = Q$. The ECDLP is of the hardest kind currently known; the only attacks that work in $(E(\mathbb{F}_q), +)$ are the ones for general groups.

Advantages of this group:

- The DLP is hard (in general);
- It is easy to describe and perform the operation.

Theorem 1.10.22. *If E is an elliptic curve over $\mathbb{F}_q$, then $E(\mathbb{F}_q) \cong \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$, where d_1 divides d_2.*

Remarks 1.10.23.

(1) *Possibly it may happen that $d_1 = 1$, i. e. $E(\mathbb{F}_q) \cong \mathbb{Z}_{d_1}$.*

(2) *Unfortunately the isomorphism is not effectively computable (similarly as in the case of $\mathbb{F}_q^* \cong \mathbb{Z}_{q-1}$).*

But what about the size of $E(\mathbb{F}_q)$?

Theorem 1.10.24 (Hasse). *If E is an elliptic curve defined over $\mathbb{F}_q$, then*

$$|E(\mathbb{F}_q)| = q + 1 - t, \qquad \text{where } |t| \leq 2\sqrt{q}.$$

Remark 1.10.25. *Consider the simplified Weierstrass equation $y^2 = x^3 + ax + b$. For every $x \in \mathbb{F}_q$, we have two points if $x^3 + ax + b$ is a quadratic residue in $\mathbb{F}_q^*$, one point if it is zero, and no points if it is a quadratic nonresidue.*

In average we expect a quadratic residue in "half of the times", thus q points is reasonable. Thus $|E(\mathbb{F}_q)| \approx q + 1$, where the $+1$ is for the point at infinity, sounds perfectly fine.

Example 1.10.26. *Consider $E : y^2 = x^3 + x + 1$ over $\mathbb{F}_{23}$. We have $\Delta = 4a^3 + 27b^2 = 4 + 4 \neq 0$, and thus E is an elliptic curve. By Hasse, we know $|E(\mathbb{F}_{23}) - 24| \leq \lfloor 2\sqrt{23} \rfloor = 9$.*

We now enumerate all the points: $\mathcal{O}$, $(0,1)$, $(0,22)$, $(1,7)$, $(1,16)$, $(3,10)$, $(3,13)$, $(4,0)$, $(5,4)$, $(5,19)$, $(6,4)$, $(6,19)$, $(7,11)$, $(7,12)$, $(9,7)$, $(9,16)$, $(11,3)$, $(11,20)$, $(12,4)$, $(12,19)$, $(13,7)$, $(13,16)$, $(17,3)$, $(17,20)$, $(18,3)$, $(18,20)$, $(19,5)$, $(19,18)$, thus we have 28 points on the curve!

Next we want to add the points $(x_1, y_1) = (3, 0)$ and $(x_2, y_2) = (9, 7)$ using the formulas from the lemma. Let the sum be (x_3, y_3); then we get

$$\begin{aligned}
\lambda &= \frac{y_2 - y_1}{x_2 - x_1} = \frac{10 - 7}{3 - 9} = -\frac{1}{2}, \\
x_3 &= \lambda^2 - x_1 - x_2 = \left(-\frac{1}{2}\right)^2 - 3 - 9 = 6 - 12 = 17, \\
y_3 &= -y_1 + \lambda(x_1 - x_3) = -10 + \left(-\frac{1}{2}\right)(3 - 17) = 20;
\end{aligned}$$

thus

$$(3, 0) + (9, 7) = (17, 20) \qquad \text{in } E(\mathbb{F}_{23}).$$

Let $(x, y) = (3, 10)$. We want to compute $2 \cdot (3, 10) = (x_4, y_4)$. We get

$$\begin{aligned}
\lambda' &= \frac{3x_1^2 + a}{2y_1} = \frac{3 \cdot 3^2 + 1}{20} = 13, \\
x_4 &= \lambda' - 2x_1 = 7, \\
y_4 &= -y_1 + \lambda'(x_1 - x_3) = -10 + 13(3 - 7) = 12;
\end{aligned}$$

thus

$$2 \cdot (3, 10) = (7, 12).$$

1.10.6 Determining the Group Order

Let E be an elliptic curve over a finite field $\mathbb{F}$. By Hasse we know that $||E(\mathbb{F})| - q - 1| \leq 2\sqrt{|F|}$. But how to effectively compute $|E(\mathbb{F})|$?

Let $P \in E(\mathbb{F}_q)$. Recall that the order of P is defined as $\operatorname{ord} P = \min\{n \geq 1 \mid nP = \mathcal{O}\}$. By Lagrange, the order of a point divides the group order. Thus $k \cdot \operatorname{ord} P = |E(\mathbb{F}_q)|$ for some $k \in \mathbb{N}$. By Hasse we know that $|E(\mathbb{F}_q)|$ lies in an interval of length $4\sqrt{q}$, and thus the choices of k are limited.

If we can find a point $Q \in E(\mathbb{F}_q)$ such that $\operatorname{ord} Q > 4\sqrt{q}$, we are done. Otherwise both $\operatorname{ord} P$ and $\operatorname{ord} Q$ divide $|E(\mathbb{F}_q)|$, and thus $|E(\mathbb{F}_q)|$ is a multiple of $\operatorname{lcm}(\operatorname{ord} P, \operatorname{ord} Q)$ (the least common multiple of $\operatorname{ord} P$ and $\operatorname{ord} Q$). Repeating this might lead to a solution. We will later see that (and also why) this does not always works.

But first we will investigate how to calculate the order of an element. This can be done by a variation of the Baby-step Giant-step algorithm: the *Shanks-Mestre algorithm*:

Shanks-Mestre The goal is to compute $\operatorname{ord}(P)$, where $P \in E(\mathbb{F}_q)$, or to compute $|E(\mathbb{F}_q)|$.

(1) Let $Q := (q + 1)P$.

(2) Choose an $m \in \mathbb{Z}$, such that $m > q^{1/4}$. (For example $m := \lfloor q^{1/4} + 1 \rfloor$.)

(3) Compute and store jP for $j = 0, \ldots, m$. (*Baby step*)

(4) Compute $Q + k(2mP)$ for $k = -m, \ldots, m$, until there is a match $Q + k(2mP) = \pm jP$ for some j, k. (*Giant step*)

(5) Compute $M := q + 1 - 2mk \mp j$. Then $M \cdot P = \mathcal{O}$.

(6) Factor M; let $p_1, \ldots, p_r$ be the distinct prime factors.

(7) Compute $\frac{M}{p_i} \cdot P \stackrel{?}{=} \mathcal{O}$. If it is equal for one i, divide M by p_i and try the i again.

(8) Now $M = \operatorname{ord} P$.

If we want to compute $|E(\mathbb{F}_q)|$, we repeat this with randomly chosen points $P_i \in E(\mathbb{F}_q)$. Repeat this until $\operatorname{lcm}(\operatorname{ord} P_1, \ldots, \operatorname{ord} P_t)$ divides only one integer in the interval $[q+1-2\sqrt{q}, q+1+2\sqrt{q}]$; this integer is then $|E(\mathbb{F}_q)|$.

Remarks 1.10.27.

The running time is $\mathcal{O}(q^{1/4+\varepsilon})$ for any constant $\varepsilon > 0$. (To be expected, since the interval contains $4\sqrt{q}$ integers.)

This algorithm works for any group G where we have bounds on $|G|$.

How hard is factoring M in step 6? If $M \approx q$, it is not too hard to factor since for elliptic curve cryptography, the values used for q are around 160 bits.

A side note on how to find a (almost uniformly distributed) random point on E. Pick an $x \in \mathbb{F}_q$ and compute $\alpha := x^3 + ax + b$. If α is a quadratic residue, find an y such that $y^2 = \alpha$ (in most cases, there are two choices; chose one randomly). Then (x, y) is a point on E. As approximately half of the x lead to a quadratic residue α, this probabilistic algorithm will in most cases need at most two tries.

A much better (since polynomial and completely deterministic) algorithm for point counting was created by Schoof in 1985. It has complexity $\mathcal{O}(\log^8 q)$. Unfortunately it is not useful for practical computations for q's of 160 bits and more. But there do exist extensions which also work good for such large q's.

Since to understand how Schoof's algorithm works requires a much deeper insight in elliptic curves and the proof of Hasse's theorem, we will not further elaborate on the algorithm. The interested reader is encouraged to consult the literature for more information.

Examples 1.10.28.

(1) *Consider the elliptic curve $E : y^2 = x^3 + 7x + 1$ over $\mathbb{F}_{101}$. Chose $P = (1, 0)$; it has order 116 (computed using Shanks-Mestre), and thus $|E(\mathbb{F}_{101})|$ is a multiple of 116. But the only multiple of 116 in the interval $[101 + 1 - 2\sqrt{101}, 101 + 1 + 2\sqrt{101}]$ is 116 itself, and thus $|E(\mathbb{F}_{101})| = 116$. Further we know that $E(\mathbb{F}_{101})$ is cyclic and generated by P.*

(2) *The point $P = (-1, 2)$ on the curve $E : y^2 = x^3 + 7x + 12$ over $\mathbb{F}_{103}$ has order 13. By Hasse we know that $84 \leq |E(\mathbb{F}_{103})| \leq 124$. Moreover, the point $Q = (19, 0)$ has order 2. Thus $|E(\mathbb{F}_{103})|$ is divisible by $\operatorname{lcm}(13, 2) = 26$, and the only possibility is $|E(\mathbb{F}_{103})| = 4 \cdot 26 = 104$. We further know that the subgroup $\langle P, Q \rangle$ has index 4.*

(3) *Consider the curve $E : y^2 = x^3 + 2$ over $\mathbb{F}_7$. Let $N := |E(\mathbb{F}_7)|$. By Hasse, $N \in [1, 13]$. Pick $P = (0, 3)$; then $2P = (0, 4)$, and $3P = \mathcal{O}$, thus $\operatorname{ord} P = 3$. So N is a multiple of 3, and thus we are left with $N \in \{3, 6, 9, 12\}$. Pick $Q = (3, 1)$; again $\operatorname{ord} Q = 3$. This looks like we have no further information about N. But $Q \notin \langle P \rangle$, and since $E(\mathbb{F}_7)$ is Abelian it follows that $\langle P, Q \rangle$ has 9 elements. Thus 9 divides N, and so we get that $N = 9$ and $E(\mathbb{F}_7) = \langle P, Q \rangle \cong \mathbb{Z}_3 \times \mathbb{Z}_3$. This shows that we were not lucky, since every element of $E(\mathbb{F}_7)$ except $\mathcal{O}$ has order 3.*

Remarks 1.10.29.

(1) *We know $E(\mathbb{F}_q) \cong \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2}$, where d_1 divides d_2. Thus $\operatorname{ord} P$ divides d_2 for every $P \in E(\mathbb{F}_q)$. If $d_2 > 4\sqrt{q}$, then by Hasse we get $|E(\mathbb{F}_q)|$ if we find d_2. (Because there is only one multiple of d_2 in the interval given by Hasse.) In the applications this will always be the case.*

However, as we saw in the third example, looking at $m = |\langle P, Q \rangle|$ gives more information than looking simply at $\ell = \operatorname{lcm}(\operatorname{ord} P, \operatorname{ord} Q)$, since ℓ divides m, which in turn divides $|E(\mathbb{F}_q)|$.

1.10.7 General Algorithms to Solve the ECDLP

These are the only algorithms known to solve the ECDLP. We work in $G = E(\mathbb{F}_q)$ with additive notation. Let $n = \operatorname{ord} P$.

(1) *Baby-step Giant-step* method:

This requires about $\mathcal{O}(\sqrt{n})$ storage and $\mathcal{O}(\sqrt{n})$ operations.

(2) *Pohlig-Hellmann algorithm*:

If $n = \prod_{i=1}^{t} p_i^{e_i}$, where $e_i \geq 1$ and the p_i's are distinct primes, the complexity of Pohlig-Hellmann is given by

$$\mathcal{O}\left(\sum_{i=1}^{t} e_i(\log n + \sqrt{p_i})\right).$$

This is good when all the p_i's are "small".

Remark 1.10.30. *Because of this algorithm we want to have a big prime p such that p divides $|E(\mathbb{F}_q)|$.*

In practice, curves with $|E(\mathbb{F}_q)|$ being prime itself, or being twice a prime are preferred.

(3) *Pollard ρ and λ method*:

The best algorithm known for elliptic curves. The complexity is $\mathcal{O}(\sqrt{n})$, and a negligible amount of storage used. First we want to recap how the Pollard ρ algorithm works. We want to find a $k \in \mathbb{Z}$ such that $kP = Q$, for $P, Q \in E(\mathbb{F}_q)$ and $Q \in \langle P \rangle$.

(1) Produce a random collection of triples (c_i, n_i, m_i) such that $c_i = n_i P + m_i Q$.

(2) Expect a collision $c_i = c_j$ but $(n_i, m_i) \neq (n_j, m_j)$ after $\approx \sqrt{n}$ steps.

(3) It is $n_i P + m_i Q = n_j P + m_j Q$, and thus $k = \frac{n_i - n_j}{m_j - m_i} \pmod{\operatorname{ord} P}$.

Example 1.10.31. *First we want to construct a "random looking" function $h : E(\mathbb{F}_q) \to \{1, 2, 3\}$. For this pick $P = (x, y) \in E(\mathbb{F}_q)$. Let $\hat{x}$ denote the integer representation of the eight least significant bits of the binary representation of x. (Thus $\hat{x} \in \{0, 1, \ldots, 255\}$.) Define $h(P)$ to be i, if $(i-1)\frac{255}{3} \leq \hat{x} < i\frac{255}{3}$. Take $S_i := h^{-1}(i)$; then we have*

$$E(\mathbb{F}_q) = S_1 \dot{\cup} S_2 \dot{\cup} S_3.$$

We use this to construct a random walk in $E(\mathbb{F}_q)$:

- *Start from a random (c_0, n_0, m_0).*
- *Generate a sequence by*

$$(c_{i+1}, n_{i+1}, m_{i+1}) = \begin{cases} (c_i + P, n_i + 1, m_i) & \text{if } c_i \in S_1, \\ (c_i + Q, n_i, m_i + 1) & \text{if } c_i \in S_2, \\ (2c_i, 2n_i, 2m_i) & \text{if } c_i \in S_3. \end{cases}$$

More complicated versions of this idea are used in the praxis.

What if $\sqrt{n}$ is too big for one machine? There is a distributed version of the Pollard ρ algorithm, called the Pollard λ algorithm. The key idea for this is to chose a sparse, random $D \in E(\mathbb{F}_q)$ such that it is easy to test whether $P \in D$ or not.

Assume we have M machines, each of them computing their own independent Pollard ρ sequence from a starting point which is the same for every machine.

If $c_i \in D$ for some c_i on some machine, it reports the triple (c_i, n_i, m_i) to a central server, which keeps track of all the triples send to it in order to find collisions.

If the sequences on two machines have one common point, this is with a high probability not in D. But from this point on the sequences follow the same way, and eventually they

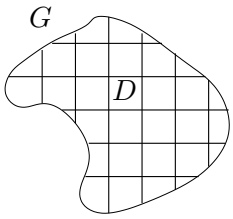


Figure 1.4: The sparse subset D of the group G

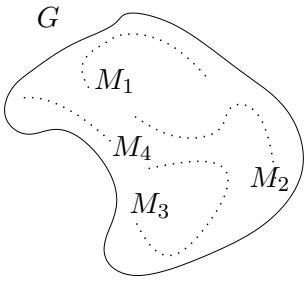


Figure 1.5: Random walks of different machines

will both hit D and get reported; in that case the central server can solve the DLP. The collision can be depicted like in picture 1.6.

This picture explains where the name "Pollard λ" comes from: the two paths look like the Greek letter λ. The complexity for M machines is $\mathcal{O}(\frac{\sqrt{n}}{M})$.

Solving the ECDLP for q of 109 bits is harder than factoring a 512 bit number. A company called Certicom, which specializes in producing crypto products based on elliptic curves set up several challanges, consisting of ECDLP's of different sizes. One of them was an ECDLP for q of 109 bits, which took over a year to solve with the help of the Pollard λ algorithm running on 10,000 machines. The software for this was developed by Chris Monico, a former student of Professor Rosenthal.

We want to give some good parameter choices for elliptic curve cryptography which are currently in use:

- $q \approx 2^{160}$, i. e. q has 160 bits;
- $q = 2^\ell$, ℓ prime, or q being a prime itself;
- $|E(\mathbb{F}_q)| \in \{p, 2p\}$, where p is prime;
- E "random" (avoiding some special curves which are "weak"; more to that later).

1.10.8 Divisors and the Weil Pairing

Rational Functions and Divisors

Definition 1.10.32. *Let $g, h \in \mathbb{F}[x, y, z]$ be two homogenous polynomials of the same degree. If $f := \frac{g}{h}$ we call f a* rational function*. If C is a curve, we say that f is* defined over C *if h does not vanishes completely on C.*

We will further also call $\frac{g}{h}$ a rational function if the degrees of g and h differ; in that case we will mean the rational function $\frac{gz^k}{hz^\ell}$ where $k, \ell \in \mathbb{N}$ are chosen minimal such that gz^k and hz^ℓ have the same degree.

Remark 1.10.33. *Recall that the zeros of a homogenous polynomial in projective space are well-defined.*

Example 1.10.34. *Let $p = x \in \mathbb{Q}[x, y, z]$. Let $Q = (x, y, z) \in \mathbb{P}^2_{\mathbb{Q}}$. If for example $Q = (1, 0, 0) = (r, 0, 0)$ for every $r \in \mathbb{Q}^*$, it doesn't makes sense to evaluate $p(Q)$, since $1 = p(1, 0, 0) = p(Q) = p(2, 0, 0) = 2$! But if $R = (a, b, c) \in \mathbb{P}^2_{\mathbb{Q}}$, then $p(a, b, c) = 0$ if and only if $p(\lambda a, \lambda b, \lambda c) = \lambda p(a, b, c) = 0$ for any $\lambda \in \mathbb{Q}^*$. Thus $p(R) = 0$ is well-defined.*

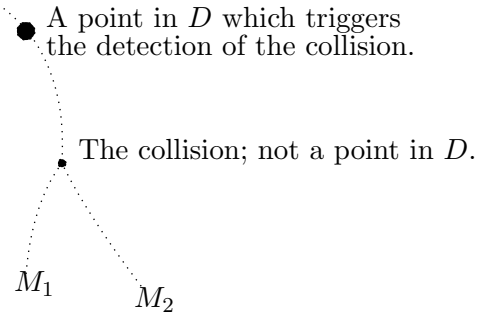


Figure 1.6: A collision and the origin of the name

Definition 1.10.35. *Let E be an elliptic curve. For each $P \in E$ define a symbol $[P]$. Denote the free Abelian group generated by the $[P]$, $P \in E$, by* $\mathrm{Div}(E)$*; these are formal sums of the form*

$$\sum_{P \in E} n_P[P], \qquad n_P \in \mathbb{Z} \text{ and } n_P = 0 \text{ for all but a finite number of } P \in E.$$

The elements of this group are called divisors *on E.*

Definition 1.10.36. *Let $f = \frac{p}{q}$ be a rational function on an elliptic curve E which is not completely vanishing on E; further assume that p and q have the same degree. Let $P_1, \ldots, P_s$ be the (distinct) zeroes of p on E with multiplicities n_i, and $Q_1, \ldots, Q_t$ the (distinct) zeroes of q on E with multiplicities m_i. Then we define the* divisor of f *as*

$$\mathrm{div}(f) := \sum_{i=1}^{s} n_i[P_i] - \sum_{i=1}^{t} m_i[Q_i].$$

Note that multiplying a polynomial with z increases the multiplicity of its intersection with $\mathcal{O}$ by one. Thus requireing that p and q have the same degree is the same than "padding" the divisor with $\mathcal{O}$'s such that if $\mathrm{div}(f) = \sum n_P[P]$*, then* $\sum n_P = 0$.

It can be shown that $\mathrm{div}(f)$ is indeed a divisor on E. Moreover, if $D = \sum n_P[P] \in \mathrm{Div}(E)$ is the divisor of a function, then $\sum n_P = 0$ and $\sum n_P P = \mathcal{O}$. The last two definitions and this statement (except that about the sum in $(E, +)$) also hold on general smooth curves.

Example 1.10.37. *Let $E : y^2 = x^3 + x + 1$ an elliptic curve over $\mathbb{F}_5$. Consider the rational function $f = \frac{x+z}{y-z}$. What is* $\mathrm{div}(f)$*?*

The zeroes of $x + z$*: If $x = -z = 0$, then the point is $\mathcal{O}$. If $x = -z \neq 0$, we have the points $(-1, \pm\sqrt{-1}, 1) = (-1, \pm 2, 1)$. Thus the zeroes of $x + z$ are*

$$[\mathcal{O}] + [(-1, 2)] + [(-1, -2)].$$

The zeroes of $x - y$*: If $x = y = 0$, this cannot be a point on the curve. Thus consider $x = y \neq 0$. In that case $1 = x^3 + x + 1$, and thus $x(x^2 + 1) = 0$. So we get $x = 0$ and $x = \pm 2$, and thus*

$$[(0, 1)] + [(0, -2)] + [(0, 2)].$$

Summing up we have

$$\mathrm{div}(f) = \big([\mathcal{O}] + [(-1, 2)] + [(-1, -2)]\big) - \big([(0, 1)] + [(0, -2)] + [(0, 2)]\big).$$

In the group $(E, +)$ we have

$$\mathcal{O} + (-1, 2) + (-1, -2) = \mathcal{O} \qquad \text{and} \qquad (0, 1) + (0, -2) + (0, 2) = \mathcal{O}.$$

Lemma 1.10.38. *Let $\alpha_i \geq 1$ and $P_i \in E \setminus \{\mathcal{O}\}$ distinct points satisfying $\sum_i \alpha_i P_i = \mathcal{O}$. Then*

$$\sum_i \alpha_i [P_i] - \sum_i \alpha_i [\mathcal{O}]$$

is the divisor of a rational function.

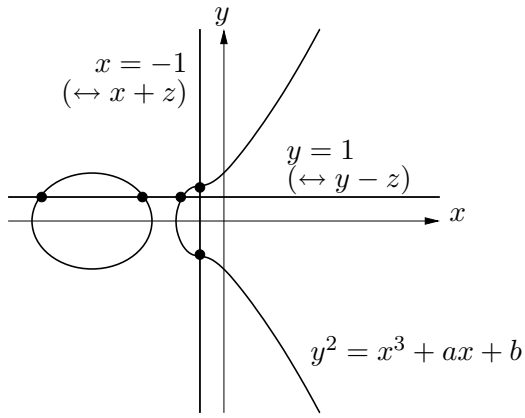


Figure 1.7: The divisor for $f = \frac{x+z}{y-z}$

Proof. We will only proof this for $\alpha_i = 1$. We will proceed by induction on i. For $i = 0$, consider the rational function $f = 1$; its divisor is 0. It is not possible that $i = 1$, since $P_1 \neq \mathcal{O}$.

If $i = 2$, then $P_1 + P_2 = \mathcal{O}$, and thus $P_2 = -P_1$, and we have $P_1 \neq -P_1$. Consider $f = x - x_P y$ if $P_1 = (x_P, y_P)$; then

$$\operatorname{div}(f) = [P_1] + [P_2] - 2[\mathcal{O}].$$

If $i = 3$, then $P_1 + P_2 + P_3 = \mathcal{O}$, and thus P_1, P_2, P_3 lie on a line which has an equation $L(x, y, z) = 0$. Then $\operatorname{div}(L) = [P_1] + [P_2] + [P_3] - 3[\mathcal{O}]$.

Now for the induction step $i - 1 \to i$. Let $P_1 + \cdots + P_i = \mathcal{O}$, and define $Q := P_1 + \cdots + P_{i-2}$. Then $Q + P_{i-1} + P_i = \mathcal{O}$, and thus there is a line equation $L(x, y, z) = 0$ such that $\operatorname{div}(L) = [Q] + [P_{i-1}] + [P_i] - 3[\mathcal{O}]$. By induction consider $P_1 + \cdots + P_{i-2} + (-Q) = \mathcal{O}$; find a rational function g such that $\operatorname{div}(g) = [P_1] + \cdots + [P_{i-2}] + [-Q] - (i-1)[\mathcal{O}]$. Let M be the equation of the vertical line through Q and $-Q$; then $\operatorname{div}(M) = [Q] + [-Q] - 2[\mathcal{O}]$. Let $f := \frac{gL}{M}$; then

$$\begin{aligned}\operatorname{div}(f) &= ([P_1] + \cdots + [P_{i-2}] + [-Q] - (i-1)[\mathcal{O}]) \\ &\quad + ([Q] + [P_{i-1}] + [P_i] - 3[\mathcal{O}]) \\ &\quad - ([Q] + [-Q] - 2[\mathcal{O}]) \\ &= [P_1] + \cdots + [P_i] - i[\mathcal{O}].\end{aligned}$$

□

Note that if f and g are rational functions such that $\operatorname{div}(f) = \operatorname{div}(g)$, then $f = \lambda g$ with a constant $\lambda \in \mathbb{F}^*$. This implies that the function associated to a divisor as in the above lemma is determined up to multiplication by non-zero constants.

Definition 1.10.39. *Denote the set of divisors $D \in \operatorname{Div}(E)$ such that there is a rational function f satisfying $\operatorname{div}(f) = D$, and the "zero" divisor 0 by $\operatorname{PDiv}(E)$. The elements of $\operatorname{PDiv}(E)$ are called the* principal divisors *on E.*

The set of principal divisors is in fact a subgroup of the group of divisors of an elliptic curve.

n-Torsion Points

Definition 1.10.40. *Let n be a natural number and E an elliptic curve. Then*

$$E[n] := \{P \in E \mid nP = \mathcal{O}\}$$

*are called the n-*torsion points *of E.*

Remarks 1.10.41.

(1) *If $P \in E[n]$, then $\operatorname{ord} P$ divides n.*

(2) *If $P, Q \in E[n]$, then $P - Q \in E[n]$ as one can simply check; hence $E[n]$ is a subgroup of $(E, +)$.*

(3) *Consider the map* $\psi : E \to E$, $P \mapsto nP$. *This is a group homomorphism, as a simple calculation shows. Moreover,* $\ker \psi = E[n]$.

Lemma 1.10.42. *Let* E *be an elliptic curve over* $\mathbb{F}$, *where* $p := \operatorname{Char} \mathbb{F} > 0$. *If* $p \nmid n$, *then* $|E[n]| = n^2$.

Sketch of Proof. Consider $\varphi_n : E \to E$, $P \mapsto nP$. This is a morphism of the curve, whose kernel is $E[n]$. By induction one can show that $\varphi_n = (\frac{p}{q}, y\frac{f}{g})$ with polynomials $p, q, f, g \in \mathbb{F}[x]$, and

$$p = x^{n^2} + \text{ lower degree terms}, \qquad q = nx^{n^2-1} + \text{ lower degree terms}.$$

We have $p' \neq 0$ since p does not divides n. To find the kernel of φ_n we have to solve $p(x) = 0$, and thus we expect n^2 solutions over the algebraic closure. Using this one can conclude that $|E[n]| = n^2$. □

Example 1.10.43. *Let* $n = 2$ *and* $p = \operatorname{Char} \mathbb{F} > 2$, *and* E *an elliptic curve given by* $y^2 = f(x)$, *where* $f \in \mathbb{F}[x]$ *is monic of degree* 3. *Write* $f = \prod_{i=1}^{3}(x - \alpha_i)$. *Now we know that*

$$E[2] = \{P \in E \mid 2P = \mathcal{O}\} = \{P \in E \mid P = -P\} = \{\mathcal{O}, (\alpha_1, 0), (\alpha_2, 0), (\alpha_3, 0)\}.$$

Thus $|E[2]| = 4 = 2^2$, *and moreover we see that* $E[2] \cong \mathbb{Z}_2 \times \mathbb{Z}_2$.

In general $E[n] \cong \mathbb{Z}_n \times \mathbb{Z}_n$ if n is not divided by $\operatorname{Char} \mathbb{F}$. If n is prime, this is an easy consequence of the structure theorem of Abelian groups, since then $E[n] \cong \mathbb{Z}_{d_1} \times \mathbb{Z}_{d_2} \times \cdots \times \mathbb{Z}_{d_t}$, where $d_1 \mid d_2, \ldots, d_{t-1} \mid d_t$ and $d_t \mid n^2$. Since it cannot be that $E[n] \cong \mathbb{Z}_{n^2}$, it follows that $t = 2$ and $d_1 = d_2 = n$ since n is prime.

Theorem 1.10.44. *Let* $n \in \mathbb{N}$, $n \geq 1$, *and let* n *not be divisible with* $\operatorname{Char} \mathbb{F}$. *Then there exists a pairing*

$$e_n : E[n] \times E[n] \to \mu_n := \{x \in \overline{\mathbb{F}} \mid x^n = 1\},$$

called the Weil pairing, *such that the following properties hold:*

(1) *The map* e_n *is* bilinear, *i. e. for any* $S, S_1, S_2, T, T_1, T_2 \in E[n]$ *we have*

$$e_n(S_1 + S_2, T) = e_n(S_1, T)e_n(S_2, T) \qquad \text{and} \qquad e_n(S, T_1 + T_2) = e_n(S, T_1)e_n(S, T_2).$$

This especially implies $e_n(S, \mathcal{O}) = 1 = e_n(\mathcal{O}, T)$ *for all* $S, T \in E[n]$.

(2) *For all* $S, T \in E[n]$ *we have* $e_n(S, T) = e_n(T, S)^{-1}$.

(3) *Fix one* $T \in E[n]$. *If* $e_n(S, T) = 1$ *for all* $S \in E[n]$, *then* $T = \mathcal{O}$.

(4) *If* $\sigma \in \operatorname{Gal}(\overline{\mathbb{F}}/\mathbb{F})$, *i. e.* σ *is a field automorphism* $\sigma : \mathbb{F} \to \mathbb{F}$ *satisfying* $\sigma|_{\mathbb{F}} = \mathbf{id}_{\mathbb{F}}$, *then* $e_n(\sigma(S), \sigma(T)) = \sigma(e_n(S, T))$ *for all* $S, T \in E[n]$. *(Here as usual* $\sigma(S)$ *denotes* $(\sigma(x), \sigma(y))$ *if* $S = (x, y)$, *and* $\mathcal{O}$ *if* $S = \mathcal{O}$.*)*

(5) *If* $\mathbb{L}$ *is a field extension of* $\mathbb{F}$, *then* $E[n] \subseteq E(\mathbb{L})$ *implies* $\mu_n \subseteq \mathbb{L}$.

The subgroup μ_n *of* $\overline{\mathbb{F}}^*$ *is called the group of* n-th roots of unity.

Proof. **Construction of the Pairing:** Fix some $T \in E[n]$; we want to define $e_n(S, T)$ for every $S \in E[n]$. We know that $nT = \mathcal{O}$, and thus $nT - n\mathcal{O} = \mathcal{O}$. Choose an $T' \in E[n^2]$ such that $nT' = T$. The existence of such an T' can be seen by considering the map $E[n^2] \to E[n]$, $P \mapsto nP$; its kernel is $E[n]$, and since $\left|E[n^2]\right| = n^4 = |E[n]| \cdot |\ker(P \mapsto nP)| < \infty$ it is surjective.

Consider the following sum in E:

$$\sum_{R \in E[n]} (T' + R) - \sum_{R \in E[n]} R = \sum_{R \in E[n]} T' = n^2 T' = nT = \mathcal{O}.$$

By Lemma 1.10.38 we can find a rational function g such that

$$\operatorname{div}(g) = \sum_{R \in E[n]} [T' + R] - \sum_{R \in E[n]} [R].$$

Since $T' + R$ runs through all the points $T'' \in E[n^2]$ such that $nT'' = T$ as R varies over $E[n]$, the divisor does not depends on the choice of T'.

Since $nT - n\mathcal{O} = \mathcal{O}$, we can find a rational function f such that $\operatorname{div}(f) = n[T] - n[\mathcal{O}]$ by the same lemma. Let $\psi_n : E \to E$, $P \mapsto nP$ and consider $f \circ \psi_n$. The divisor of $f \circ \psi_n$ is

$$n \sum_{R \in E[n]} [T' + R] - n \sum_{R \in E[n]} [R],$$

and this is equal to $\operatorname{div}(g^n)$. Therefore we know that

$$f \circ \psi_n = \lambda \cdot g^n, \qquad \text{where } \lambda \in \mathbb{F}^*.$$

Let $S \in E[n]$ and pick an $P \in E$. We have

$$g(P+S)^n = \tfrac{1}{\lambda} f(\psi_n(P+S)) = \tfrac{1}{\lambda} f(nP + nS) = \tfrac{1}{\lambda} f(nP) = g(P)^n,$$

and thus

$$e_n(S,T) := \frac{g(P+S)}{g(P)} \in \mu_n.$$

What is left is to show that $e_n(S,T)$ is well-defined: consider the map $f_S : E \to \mu_n$, where $P \mapsto \frac{g(P+S)}{g(P)}$; we have to show that it is constant. We will use a topological argument here: consider the discrete topology on μ_n and the cofinite topology on E (i. e. the closed sets are finite subsets of E, together with E itself). Then f_S is continuous with respect to this map. Moreover, since E is connected, f_S must be constant.

To the reader with more background in algebraic geometry: in fact the topologies chosen are the trace topologies of the Zariski topologies on $\mathbb{P}^2(\overline{\mathbb{F}})$ and $\overline{\mathbb{F}}$; and the fact that f_S is continuous follows from that it can be written as $f_S = (\frac{p}{q}, \frac{f}{g})$ with polynomials $p, q, f, g \in \mathbb{F}[x, y]$. That E is connected follows from the fact that an elliptic curve is irreducible.

The properties: Left as homework. □

The MOV Attack MOV stands for Menezes, Okanoto and Vanstone, who came up with this attack. This is an attack based on the Weil pairing, its goal being to solve the ECDLP by reducing it to a DLP in a finite extension of $\mathbb{F}_q$.

Assume that $P, Q \in E(\mathbb{F}_q)$ are given, and $N = \operatorname{ord} P$ is coprime to the field characteristic q. Moreover assume $kP = Q$ for some $k \in \mathbb{Z}$. We want to find k.

Lemma 1.10.45. *If $Q = kP$, then $NQ = 1$ and $e_N(P,Q) = 1$.*

Proof. We have $NQ = NkP = k(NP) = k\mathcal{O} = \mathcal{O}$; thus $P, Q \in E[n]$ and we can evaluate $e_N(P,Q)$: it is

$$e_N(P,Q) = e_N(P,kP) = (e_N(P,P))^k = 1^k = 1.$$

□

The *MOV attack* works as follows:

(1) Choose an m such that $E[N]$ embeds into $E(\mathbb{F}_{q^m})$. Such an m exists since $E[N]$ is finite and $\overline{\mathbb{F}_q} = \bigcup_{\ell \geq 1} \mathbb{F}_{q^\ell}$.

(We also get $\mu_N \subseteq \mathbb{F}_{q^m}^*$ by the theorem.)

(2) Choose a random point $T \in E(\mathbb{F}_{q^m})$ and compute $M := \operatorname{ord} P$.

(3) Take $d := \gcd(M, N)$, and let $T_1 := \frac{M}{d} \cdot T$. Thus $d = \operatorname{ord} T_1$ divides N, and so $T_1 \in E[N]$.

(4) Set $\xi_1 := e_N(P, T_1)$ and $\xi_2 := e_N(Q, T_1)$. Then $\xi_1, \xi_2 \in \mu_N \subseteq \mathbb{F}_{q^m}^*$. Moreover $\xi_1, \xi_2 \in \mu_d$, since $\xi_1^d = e_N(P,T_1)^d = e_N(P, dT_1) = e_N(P, \mathcal{O}) = 1$ and similarly $\xi_2^d = 1$.

(5) Since we have

$$\xi_2 = e_N(Q, T_1) = e_N(kP, T_1) = (e_N(P, T_1))^k = \xi_1^k,$$

by solving the DLP $\xi_1^k = \xi_2$ in $\mu_d \subseteq \mathbb{F}_{q^m}^*$ we get $k \mod d$.

(6) Repeat from step 2 until the least common multiple of the d's is N. In this case, one can use the Chinese Remainder Theorem to recover k.

Remark 1.10.46. *We reduce one DLP in $E(\mathbb{F}_q)$ to several DLP's in $\mathbb{F}_{q^m}$. Since $|\mathbb{F}_{q^m}|$ grows exponentially when m grows, this method gets useless if m is too large.*

There is a family of curves for which m is bounded by six; these are the supersingular curves. *In case $|E(\mathbb{F}_q)| = q + 1$, even worse m is bounded by two. This is why supersingular curves are avoided when a hard DLP is required. (Please note that being supersingular has nothing to do with being singular!)*

1.11 Alternative Public-Key Systems

1.11.1 Rabin System (1981)

Assume that p and q are primes satisfying $3 \leq p < q$, and let $n := pq$. Let b and c be integers. What are the possible solutions $x \in \mathbb{Z}_n$ of

$$x^2 + bx + c \pmod{n}?$$

Remark 1.11.1. *Since $p, q > 2$ we have that 2 is invertible in $\mathbb{Z}_n$. Thus we can write*

$$x^2 + bx + c = (x + \tfrac{b}{2})^2 - (\frac{b^2}{4} - c),$$

and hence solving $x^2 + bx + c = 0$ is equivalent to solving $z^2 - \alpha = 0$, where $z = x + \frac{b}{2}$ and $\alpha = \frac{b}{4} - c$.

Lemma 1.11.2. *The equation $z^2 - \alpha \equiv 0 \pmod{n}$ has at most four solutions in $\mathbb{Z}_n$.*

Proof. Modulo p it has at most two solutions, since $\mathbb{Z}_p$ is a field, and the same holds for modulo q. By the Chinese Remainder Theorem, these respect to a maximum of four solutions modulo $n = pq$. □

Lemma 1.11.3. *Assume $z^2 - \alpha = 0$ has solutions $\pm s \pmod{p}$ and $\pm t \pmod{q}$. Let u, v be integers such that $up + vq = 1$ (Bezóut equation). Then the general solution of $z^2 - \alpha = 0 \pmod{n}$ is given by*

$$\pm t \cdot up \pm s \cdot vq.$$

Proof. This follows from the Chinese Remainder Theorem and the fact that

$$up \equiv 1 \pmod{q},\ vq \equiv 0 \pmod{q} \qquad \text{and} \qquad up \equiv 0 \pmod{p},\ vq \equiv 1 \equiv q.$$

□

Remark 1.11.4. *It is possible that $z^2 - \alpha$ has zero, one, two or four solutions. Three solutions are not possible; as the number of solutions modulo $n = pq$ equals the number of solutions modulo p times the number of solutions modulo q.*

Remark 1.11.5. *Consider $p = 3$, $q = 5$ and thus $n = 15$. In $\mathbb{Z}_{15}$ we have*

x	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14
x^2	0	1	4	9	1	10	6	4	4	6	10	1	9	4	1

The equation $z^2 = 0$ has one solution, the equation $z^2 = 1$ four solutions, the equation $z^2 = 6$ has two solutions and the equation $z^2 = 2$ has no solutions modulo 15.

In order to compute solutions of $x^2 + bx + c = 0$ it is enough to compute solutions of $z^2 - \alpha = 0$ in the finite fields $\mathbb{F}_p$ and $\mathbb{F}_q$. In a general field $\mathbb{F}_r$, one has a probabilistic polynomial time algorithm called *Shank's algorithm* to achieve this task. But there is a special situation where solving $z^2 - \alpha = 0 \pmod{p}$ is easy:

Lemma 1.11.6. *Assume p is a prime satisfying $p \equiv 3 \pmod{4}$. Assume that $z^2 - \alpha = 0$ has a solution modulo p. Then the solutions are given by*

$$z_{1,2} = \pm \alpha^{\frac{p+1}{4}}.$$

Proof. Assume $z^2 - \alpha = 0$ has a solution in $\mathbb{F}_p$. This is equivalent to $\left(\frac{\alpha}{p}\right) = 1$, which by Euler is equivalent to $\alpha^{\frac{p-1}{2}} \equiv 1 \pmod{p}$. This again is equal to

$$\left(\pm \alpha^{\frac{p+1}{4}}\right)^2 = \alpha^{\frac{p-1}{2}} \alpha \equiv \alpha \pmod{p}.$$

□

Lemma 1.11.7. *Assume $n = pq$ with distinct primes p and q, and assume one knows four different solutions $\alpha_1, \ldots, \alpha_4$ of the equation $z^2 - \alpha \equiv 0 \pmod{n}$. Then the facorization of n can be revealed in polynomial time.*

Proof. We know that $\alpha_1, \ldots, \alpha_4$ are of the form $\pm svq \pm tup$, where $up + vq = 1$ with $u, v \in \mathbb{Z}$. We can assume that $\alpha_1 = -\alpha_4$ and $\alpha_2 = -\alpha_3$. Then

$$0 = \alpha_1^2 - \alpha_2^2 = (\alpha_1 - \alpha_2)(\alpha_1 + \alpha_2).$$

Since $\alpha_1 \neq \pm\alpha_2$ we know that $\alpha_1 - \alpha_2 \not\equiv 0 \not\equiv \alpha_1 + \alpha_2 \pmod{n}$, and thus $\gcd(\alpha_1 - \alpha_2, n), \gcd(\alpha_1 + \alpha_2, n) \in \{p, q\}$. □

The Rabin System Alice choses $p, q \geq 10^{100}$, satisfying $p \equiv q \equiv 3 \pmod{4}$ and $p \neq q$. Let $n := pq$. The *public information* is n and a randomly chosen $b \in \mathbb{Z}_n$. The *private information* is the facorization $n = pq$. *Encryption* is done by the function

$$\varphi : \mathbb{Z}_n \to \mathbb{Z}_n, x \mapsto -x^2 - bx = c.$$

For *decryption*, find solutions of $x^2 + bx + c \equiv 0 \pmod{n}$. For this solve $z^2 - \alpha = 0$ where $z = x + \frac{b}{2}$ and $\alpha = \frac{b^2}{4} - c$, by using $z \equiv \alpha^{\frac{p+1}{4}} \pmod{p}$ and $z \equiv \alpha^{\frac{q+1}{4}} \pmod{q}$.

The *costs* for this are one multiplication for encryption, and $\mathcal{O}(\log^3 n)$ bit operations for decryption.

Comparism to RSA System

- Contrary to RSA we can *prove* that breaking this the system is equivalent to factoring n.
- The encryption and decryption complexity is similar, assuming for RSA an encryption exponent like $e = 2^{16} + 1$ is chosen:

	Encryption	Decryption
Rabin	2 muls	$\mathcal{O}(\log^3 n)$ bit ops
RSA	17 muls	$\mathcal{O}(\log^3 n)$ bit ops

- But which of the four solutions of $x^2 + bx + c = 0$ was the sent message?! The Rabin scheme has the disadvantage that there are up to four square roots of α, and thus the original message cannot be recovered completely.

  This can (partially) be solved by either appending a check sum to the message, or by making an agreement of for example setting the last 30 bits of the message to zero. Unfortunately both of these methods give away information and may help an attacker.

1.11.2 The Merkle-Hellman Knapsack System

Definition 1.11.8. *Given positive integers $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$, and two positive integers s and t, the* knapsack problem *asks:*

Determine if there i a subset $S \subseteq \{1, \ldots, n\}$ such that

$$\sum_{i \in S} a_i \leq s \qquad \text{and} \sum_{i \in S} b_i \geq t.$$

A special case of this is the subset sum problem, *where $a_i = b_i$ for every i and $s = t$. In that case $\sum_{i=1}^n x_i a_i = s$ must be fulfilled for a vector $(x_i)_i \in \{0, 1\}^n$, where again $a_1, \ldots, a_n$ and s are positive integers.*

Remark 1.11.9. *Both problems are NP-complete.*

The subset sum problem is NP-complete, but certain instances of it can easily be solved in polynomial time:

Definition 1.11.10. *A sequence $a_1, a_2, a_3, \ldots$ of positive integers is called a* superincreasing set *if and only if*

$$a_j > \sum_{i=1}^{j-1} a_i \qquad \text{for all } j = 2, 3, \ldots.$$

If $a_1, \ldots, a_n$ is a superincreasing set, the subset sum problem has an easy algorithm: Construct inductively $x_n, x_{n-1}, \ldots, x_1 \in \{0, 1\}$ such that $\sum_{i=1}^n x_i a_i = s$ as follows: If $x_{j+1}, \ldots, x_n$ are chosen, check whether $s < a_j + \sum_{i=j+1}^n x_i a_i$; if that is the case, let $x_j := 0$, and otherwise $x_i := 1$.

One can easily see that this works; in fact the solution, if it exists, is unique. If it does not exists, the algorithm will terminate with $\sum_{i=1}^n x_i a_i < s$. The complexity of this algorithm is linear; in case the a_i are given in random order and have to be sorted first, the complexity increases[9] to $\mathcal{O}(n \log n)$.

The Merkle-Hellman System (1978)

- First, choose a superincreasing sequence $a_1, \ldots, a_n$.
- Choose an $m \in \mathbb{N}$ such that $m > \sum_{i=1}^n a_i$.
- Choose a random $s \in \mathbb{Z}_m^*$ and a permutation $\pi \in S_n$.
- *Publish* $b_i := s a_{\pi(i)} \mod m$ for $i = 1, \ldots, n$ and n.
- The *private information* is π, $a_1, \ldots, a_n$ and s. In fact, m can both be made public or private.
- *Encryption* is done by $\varphi : \{0,1\}^n \to \mathbb{Z}_m$, $(x_i)_i \mapsto \sum_{i=1}^n x_i b_i = c$.
- *Decryption* is done by computing $s^{-1}c = \sum_{i=1}^n x_i a_{\pi(i)}$, and thus x_i can easily be computed since the a_i are superincreasing. (And in fact we even know how they are ordered.)

The system looked very attractive in 1980:

- The subset sum problem was known to be NP-complete (that is not the case for factoring).
- It is suggested to take $n = 100$, $a_1 \approx 2^{100}$, …, $a_n = a_{100} \approx 2^{200}$ and hence $m \approx 2^{200}$. In this case, the public key is around 20 kByte. Encryption and decryption require around 100 additions of 200 bit numbers. In comparism, RSA requires around 1000 multiplications of 1000 bit numbers for decryption.

Early it was recognized that the a_i's should not be chosen too simple, e. g. $a_1 \approx 2^{100}$ and $a_{j+1} = 2a_j$, $j = 1, \ldots, n-1$ would be a bad choice: in this case, compute all possible differences $|b_i - b_j|$ for all $1 \le i < j \le n$. Sometimes $a_{j+1} = 2a_j$ and sometimes $a_{j+1} = 2a_j - m$ in $\mathbb{Z}$. Among the $\binom{n}{2}$ differences many times the value m appears.

In 1984 Adi Shamir found a polynomial time algorithm to solve the Merkle-Hellman problem. The idea of Shamir is based on the following observation:

As soon as an attacker can find a modulus $\tilde{m}$, a permutation $\tilde{\pi} \in S_n$ and a factor $u \in \mathbb{Z}_{\tilde{m}}^*$ such that the $\tilde{a}_i := u b_{\tilde{\pi}(i)} \mod \tilde{m}$ form a superincreasing set, then the subset sum problem can be solved as well. (Recall that the problem was constructed that there is exactly one solution.)

In 1985 Lagarias and Odlyzko showed how the Merkle-Hellman problem can be solved by searching for a shortest vector in a lattice. If the reader is new to the subject of lattices, he is recommended to first read the beginning of the next section, section 1.12.

Assume that $b_1, \ldots, b_n \in \mathbb{Z}$ is the public key, and $c = \sum_{i=1}^n x_i b_i$ the cipher. Let $N > \frac{1}{2}\sqrt{n}$ and consider the $(n+1)$-dimensional lattice generated by the matrix

$$M := \left(\begin{array}{cccc|c} 1 & 0 & \cdots & 0 & Nb_1 \\ 0 & \ddots & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & \cdots & 0 & 1 & Nb_n \\ \hline \frac{1}{2} & \cdots & \cdots & \frac{1}{2} & Nc \end{array}\right) \in \mathbb{R}^{(n+1)\times(n+1)}$$

[9] Sorting can be done in $\mathcal{O}(n \log n)$, for example by using the quick sort algorithm. Moreover it can be shown that in general a sorting algorithm cannot be more efficient than this, so in fact this complexity bound is strong.

A solution $\sum_{i=1}^{n} x_i b_i = c$ corresponds to the lattice vector

$$(x_1, \ldots, x_n, -1)M = (\pm\tfrac{1}{2}, \ldots, \pm\tfrac{1}{2}, 0).$$

It follows that the norm of this vector is $\frac{\sqrt{n}}{2}$. Note that the determinant of the lattice is

$$\det M = N\left(c - \frac{1}{n}\sum_{i=1}^{n} b_n\right).$$

Lagarias and Odlyzko showed that for suitable N the lattice has a density low enough such that the LLL algorithm will give the shortest vector. (See the end in Section 1.12.)

Remark 1.11.11. *After Merkle-Hellman new Knapsack type systems and also new attacks were discovered. An example for a new system is the Multiple Merkle-Hellman system: Given $a_1, \ldots, a_n$ a superincreasing set, choose different moduli $m_1, \ldots, m_r$ and different factors $u_1, \ldots, u_r$. Compute $b_i := \tilde{b}_{i,r}$ iteratively as $\tilde{b}_{i,0} := a_i$, $\tilde{b}_{i,j} := u_j \tilde{b}_{i,j-1} \mod m_j$ for $j = 1, \ldots, r$. This looks more complicated, but this can also be attacked with a generalized Lagarias-Odlyzko attack. The only Knapsack type system which is not (yet) broken is the Chov-Rivest system.*

1.11.3 Polly-Cracker

The basic idea is as follows:

Give polynomials $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ over a finite field $\mathbb{F}$, consider the ideal $I := \langle f_1, \ldots, f_m \rangle$. Assume it is not feasible to find any $p \in K^n$ such that

$$p \in V_K(I) = \{q \in K^n \mid f(q) = 0 \text{ for every } f \in I\} = \{q \in K^n \mid f_i(q) = 0 \text{ for } i = 1, \ldots, m\},$$

where K is a field extension of $\mathbb{F}$. Assume that the designer knows such an $p \in V_K(I)$. For example, the designer could start by chosing a random $p \in \mathbb{F}^n$ and random polynomials $\tilde{f}_i \in \mathbb{F}[x_1, \ldots, x_n]$, and then computing $f_i := \tilde{f}_i - \tilde{f}_i(p)$.

- The *public key* is the polynomial ring $\mathbb{F}[x_1, \ldots, x_n]$ together with the polynomials $f_1, \ldots, f_m$.
- The *private key* is the extension field K and the common zero $p \in K^n$.
- For *encrypting* a message $m \in \mathbb{F}$, choose random $q_1, \ldots, q_m \in \mathbb{F}[x_1, \ldots, x_n]$ and compute $c = c(m) = m + \sum_{i=1}^{m} f_i q_i$.
- For *decrypting* a cipher $c \in \mathbb{F}[x_1, \ldots, x_n]$, simply compute

$$c(p) = \left(m + \sum_{i=1}^{m} f_i q_i\right)(p) = m + \sum_{i=1}^{m} f_i(p) q_i(p) = m + \sum_{i=1}^{m} 0 \cdot q_i(p) = m.$$

Note 1.11.12.

(1) *Decryption is easy: just compute $m = c(p)$.*

(2) *If Eve knows any other solution $q \in V_{\overline{\mathbb{F}}}(I)$, she also can decode m by $m = c(q)$.*

Thus it should not be practical to find any solution in $V(I)$.

Definition 1.11.13. *An ideal $I \subseteq \mathbb{F}[x_1, \ldots, x_n]$ is called* zero-dimensional *if $V_{\overline{\mathbb{F}}}(I)$ is a finite set.*

Preferably Alice works with a zero-dimensional ideal.

Lemma 1.11.14. *Let $I \subseteq \mathbb{F}[x_1, \ldots, x_n]$ be an ideal. Then I is zero-dimensional if and only if $\mathbb{F}[x_1, \ldots, x_n]/I$ is a finite-dimensional $\mathbb{F}$-vector space.*

Question 1.11.15. *How hard is it to compute points in $V(I)$?*

Example 1.11.16. *Assume the ideal is generated by linear polynomials* $f_i = \sum_{j=1}^{n} a_{ij}x_j - b_i$, *where* $a_{ij}, b_i \in \mathbb{F}$. *In this case one has to solve the linear system* $Ax = b$, *where* $A = (a_{ij})$, $b = (b_i)_i$, *i. e. one uses Gauss elimination.*

For zero-dimensional ideals I, a "diagonalization process" is also possible transforming $f_1, \ldots, f_s$, where $\langle f_1, \ldots, f_s\rangle = I$, to $g_1, \ldots, g_n$, such that $\langle g_1, \ldots, g_n\rangle = I$ and

$$g_1 \in \mathbb{F}[x_1], \quad g_2 \in \mathbb{F}[x_1, x_2], \quad \ldots \quad g_n \in \mathbb{F}[x_1, \ldots, x_n].$$

Then solving this equations $f_1 = 0, \ldots, f_s = 0$ is equivalent to solving at least n polynomial equations in one variable: first find all x_1 such that $g_1(x_1) = 0$. Then plug in these values of x_1 into g_2, and solve $g_2(x_2) = 0$ for each of them, and so on.

We want to give a little description of the process behind finding such g_i. We first introduce a *monomial order* on $\mathbb{F}[x_1, \ldots, x_n]$; this is an total order[10] on the set of monomials $\{x^\alpha := \prod_{i=1}^{n} x_i^{\alpha_i} \mid \alpha \in \mathbb{N}^n\}$, such that it is compatible[11] with multiplication $x^\alpha x^\beta = x^{\alpha+\beta}$ and the set is well-ordered[12]. Based on this one can define the *leading term* of an polynomial, and can give a generalization of the Euclidean algorithm, the *division algorithm*:

Let $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$ and $f \in \mathbb{F}[x_1, \ldots, x_n]$. One has a "division with remainder" $f = \sum_{i=1}^{m} u_i f_i + r$, where $u_i, r \in \mathbb{F}[x_1, \ldots, x_n]$ and no term x^α appearing in r is divisible by any leading term of f_i, $1 \le i \le m$. Such a representation (with even more properties) can be computed by iteratively checking whether the leading term of any of the f_i's divides the leading term of f; if it does, let $f \leftarrow f - \lambda x^\alpha f_i$ where λx^α is a monomial such that the leading terms cancel away; if none of the f_i's satisfy this, subtract the leading term from f and add it to r. Repeat this as long $f \neq 0$.

Definition 1.11.17. *A generating set* $g_1, \ldots, g_m$ *of an ideal* $I \subseteq \mathbb{F}[x_1, \ldots, x_n]$ *is called a* Gröbner-basis *of* I *if for all* $f \in I$ *there exists an integer* i *such that the leading term of* g_i *divides the leading term of* f.

Remarks 1.11.18. *Some consequences of this definition:*

(1) *If* $g_1, \ldots, g_m$ *is a Gröbner-basis of* I, *then the remainder* r *in the division of* f *by* $g_1, \ldots, g_m$ *is unique.*

(2) *The quotient* $R = \mathbb{F}[x_1, \ldots, x_n]/I$ *has a representative system of reduced remainders:*

Given $f + I$, *calculate* $f = \sum u_i g_i + r$ *by the division algorithm. Then* $f + I = r + I$, *and* r *is unique.*

Note that if r_1, r_2 *are reduced representatives of* $r_1 + I$ *and* $r_2 + I$, *then* $r_1 + r_2$ *is reduced again, hence* $(r_1 + r_2) + I = (r_1 + I) + (r_2 + I)$ *has the representative* $r_1 + r_2$.

(3) *In fact it can be shown that any set of polynomials* $g_1, \ldots, g_n \in I$ *satisfying that if the leading terms of the* g_i *divide the leading term of any polynomial in* I, *then* $g_1, \ldots, g_n$ *are a generating set of* I.

Generalized Polly-Cracker

- *Public* are $f_1, \ldots, f_m \in \mathbb{F}[x_1, \ldots, x_n]$, and reduced elements $r_1, \ldots, r_t$ for $R = \mathbb{F}[x_1, \ldots, x_n]/I$, where $I = \langle f_1, \ldots, f_m\rangle$ and the r_i's are linearly independent over $\mathbb{F}$.
- *Private* is the Gröbner-basis $g_1, \ldots, g_s$ of I. (If I is zero-dimensional, then $s = n$.)

[10] Let $\le$ be a relation on a set X. Then $\le$ is called an *order* on X if for every $a, b, c \in X$ we have

(i) $a \le a$ (we say $\le$ is reflexive),

(ii) $a \le b$ and $b \le a$ implies $a = b$ (we say $\le$ is antisymmetric),

(iii) $a \le b$ and $b \le c$ implies $a \le c$ (we say $\le$ is transitive).

If moreover for every pair $a, b \in X$ at least one of $a \le b$ and $b \le a$ holds, then $\le$ is called a *total order*.

[11] An order $\le$ on a group G is said to be *compatible* with the group operation $+$ if for any $a, b, c \in G$ satisfying $a \le b$ we have $a + c \le b + c$.

[12] A total ordered set is called *well-ordered* if every non-empty subset contains a smallest element. A well-known example is $(\mathbb{N}, \le)$, and a counterexample is $(\mathbb{Z}, \le)$.

- To *encrypt*, let $m \in \mathbb{F}^t$ and define

$$c(m) := \sum m_i r_i + \sum h_i f_i, \qquad \text{where the } h_i \in \mathbb{F}[x_1, \ldots, x_n] \text{ are random.}$$

- To *decrypt*, compute $\sum m_i r_i$ as the reduction of $c(m)$ by the Gröbner-basis. From this one can then further reconstruct the r_i.

1.11.4 McEliece Crypto System (1978)

This system is based on coding theory; thus we first give a small background in that area.

1.11.4.1 A Small Background in Coding Theory

Let $\mathbb{F}$ be a finite field.

Definition 1.11.19. *A subspace $C \subseteq \mathbb{F}^n$ is called a* linear code. *If $k = \dim_{\mathbb{F}} C$, then C is called an $[n, k]$* linear code. *If $\mathbb{F} = \mathbb{F}_2$, we also call C a* binary code.

Example 1.11.20. *The ISBN code is the subspace*

$$C := \{(x_i)_i \in \mathbb{F}_{11}^{10} \mid \sum_{i=1}^{10} i x_i = 0\}.$$

In practice, for $x_1, \ldots, x_9$ only the digits 0 to 9 are used, and the tenth value x_{10} is calculated by

$$x_{10} = \sum_{i=1}^{9} i x_i.$$

For x_{10} it might happen that $x_{10} = 10$; in that case, an "X" is used in the representation if an ISBN as a ten digit number.

Definition 1.11.21. *Given $x, y \in \mathbb{F}^n$, one defines the* Hamming distance *as*

$$Ham(x, y) := |\{i \mid x_i \neq y_i\}|.$$

The distance *of a code $C \subseteq \mathbb{F}^n$ is defined as*

$$d(C) := \min\{Ham(x, y) \mid x, y \in C,\ x \neq y\}.$$

Remark 1.11.22. *The function $Ham : \mathbb{F}^n \times \mathbb{F}^n \to \mathbb{N}$ is a* metric*:*

(i) *for every $x, y \in \mathbb{F}^n$, we have $Ham(x, y) \geq 0$, and $Ham(x, y) = 0$ if and only if $x = y$;*

(ii) *for every $x, y \in \mathbb{F}^n$ we have $Ham(x, y) = Ham(y, x)$;*

(iii) *for every $x, y, z \in \mathbb{F}^n$ we have $Ham(x, z) \leq Ham(x, y) + Ham(y, z)$.*

Moreover, this metric is translation invariant*: for every $x, y, z \in \mathbb{F}^n$, we have $Ham(x, y) = Ham(x + z, y + z)$, and thus especially $Ham(x, y) = Ham(x - y, 0)$.*

Lemma 1.11.23. *Let C be a linear code. Then up to $(d(C) - 1)$ errors can be detected and up to $\left\lfloor \frac{d(C)-1}{2} \right\rfloor$ errors can be corrected, in the following sense:*

If someone sends $x \in \mathbb{F}^n$, and another one receives $x + e$, where $e \in \mathbb{F}^n$ is an error vector*, one can correctly recover x from $x + e$ if $Ham(e, 0) \leq \left\lfloor \frac{d(C)-1}{2} \right\rfloor$ by chosing the unique $\tilde{x} \in C$ such that $Ham(x + e, \tilde{x}) \leq \left\lfloor \frac{d(C)-1}{2} \right\rfloor$, and one can detect that $x + e \neq x$ without knowing what e or x is if $Ham(e, 0) \leq d(C) - 1$.*

Example 1.11.24. *Assume we want to encode the four directions "west", "north", "east" and "south". One possible code would be*

$$\textit{west} \mapsto 00, \quad \textit{north} \mapsto 01, \quad \textit{east} \mapsto 10, \quad \textit{south} \mapsto 11.$$

This scheme has distance 1*. By adding a check digit*

$$\textit{west} \mapsto 000, \quad \textit{north} \mapsto 011, \quad \textit{east} \mapsto 101, \quad \textit{south} \mapsto 110,$$

one gets a scheme with distance 2*. The following scheme has distance* 3*:*

$$\textit{west} \mapsto 00000, \quad \textit{north} \mapsto 01101, \quad \textit{east} \mapsto 10110, \quad \textit{south} \mapsto 11111.$$

One can easily check that they are all binary linear codes of dimension 2*.*

Remark 1.11.25. *The last code $C \subseteq \mathbb{F}_2^5$ is a $[5, 2]$ linear code. Indeed, one has that*

$$C = \mathrm{rowsp}_{\mathbb{F}_2} \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \end{pmatrix}.$$

Proof of the lemma. Assume that $Ham(e, 0) \leq d(C) - 1$, when x was sent and $x + e \neq x$ was received. Now $x + e \notin C$, since otherwise $Ham(0, e) = Ham(x, x + e) \leq d(C) - 1$ and $x \neq x + e$, contradicting the definition of $d(C)$.

Now assume that the receiver knows that $Ham(e, 0) \leq t := \left\lfloor \frac{d(C)-1}{2} \right\rfloor$. Then x is the unique $\tilde{x} \in C$ satisfying $Ham(x+e, \tilde{x}) \leq t$, since if $\tilde{x}$ fulfills $Ham(\tilde{x}, x+e) \leq t$ we get $Ham(x, \tilde{x}) \leq Ham(x, x + e) + Ham(x + e, \tilde{x}) \leq t + t < d(C)$, and thus it must be that $\tilde{x} = x$. □

One can visualize this as follows: define the *(closed) ball*

$$B_\varepsilon(x) := \{y \in \mathbb{F}^n \mid Ham(x, y) \leq \varepsilon\}, \qquad \varepsilon > 0,\ x \in \mathbb{F}^n.$$

Then if we have two balls around different codewords of C with radius $\frac{d(C)-1}{2}$, the balls do not intersect.

Definition 1.11.26. *Given an $[n, k]$ linear code $C \subseteq \mathbb{F}^n$, an $k \times n$-matrix $G \in \mathbb{F}^{k \times n}$ is called a* generator matrix *of C if*

$$C = \mathrm{rowsp}_{\mathbb{F}} G,$$

and an $(n - k) \times n$-matrix $H \in \mathbb{F}^{(n-k)\times n}$ is called a parity check matrix *if*

$$C = \ker H = \{x \in \mathbb{F}^n \mid Hx^t = 0\}.$$

Example 1.11.27. *The matrix*

$$H = \big(1, 2, 3, \ldots, 9, 10\big) \in \mathbb{F}_{11}^{1\times 10}$$

is a parity check matrix for the ISBN code, and the matrix

$$G = \left(\begin{array}{ccc|c} 1 & & 0 & 1 \\ & \ddots & & \vdots \\ 0 & & 1 & 10 \end{array}\right)$$

is a generator matrix for the ISBN code.

Remark 1.11.28. *We have $GH^t = 0_{k\times(n-k)}$ and $HG^t = 0_{(n-k)\times k}$.*

Back to distances The *main linear coding problem* is:

Given $\mathbb{F}$, n, k and d, find an $[n, k]$ linear code of distance at least d and with the maximum number of codewords.

For this there are some bounds:

Sphere Packing Bound Assume $d(C) = 2t + 1$. Then

$$\bigcup_{x \in C} B_t(x) \subseteq \mathbb{F}^n$$

is a disjoint union, and thus

$$\sum_{x \in C} |B_t(x)| = \left| \bigcup_{x \in C} B_t(x) \right| \leq |\mathbb{F}^n| = |\mathbb{F}|^n .$$

Now $|B_t(x)|$ does not depends on x since the code is linear and Ham is translation invariant, and thus we have the bound

$$|C| \cdot |B_t(0)| \leq |\mathbb{F}|^n .$$

Now we have that $B_t(0)$ is a disjoint union:

$$B_t(0) = \bigcup_{r=0}^{t} \{x \in C \mid Ham(x, 0) = r\}.$$

Thus

$$|B_t(0)| = \sum_{r=0}^{t} |\{x \in C \mid Ham(x, 0) = r\}| = \sum_{r=0}^{t} \binom{r}{n} (|\mathbb{F}| - 1)^r,$$

and finally we get

$$|C| \leq \frac{|\mathbb{F}|^n}{\sum_{r=0}^{t} \binom{r}{n} (|\mathbb{F}| - 1)^r}.$$

Example 1.11.29. *The largest code in $\mathbb{F}_2^5$ of distance 3 if bounded by*

$$\frac{2^5}{1 + \binom{5}{1} \cdot 1} = 5 + \frac{2}{6},$$

and thus $|C| \leq 5$ for any $[5, 2]$ code $C \subseteq \mathbb{F}_2^5$ having distance at least 3.

Singleton Bound

Lemma 1.11.30. *Assume C is an $[n, k]$ linear code with parity check matrix H. (Thus $C = \ker H$, and $H \in \mathbb{F}^{(n-k) \times n}$.) Then*

$$d(C) = \min\{i \mid \text{ there are } i \text{ linearly dependent columns of } H\}.$$

Example 1.11.31. *For the ISBN code we get $d(C) = 2$, since*

$$H = \begin{pmatrix} 1 & 2 & \cdots & 9 & 10 \end{pmatrix} \in \mathbb{F}_{11}^{1 \times 10}.$$

For the proof we will use the following notation: if H is a matrix, $H_{\bullet i}$ will denote the i-th column of H.

Proof of the lemma. It is

$$d(C) = \min\{Ham(x, y) \mid x, y \in C,\ x \neq y\} = \min\{Ham(x, 0) \mid x \in C \setminus \{0\}\}.$$

Let $x = (x_i)_i \in C \setminus \{0\}$. The $Hx^t = 0$, and thus $\sum H_{\bullet i} x_i = 0$ is a linear combination of the i-th columns of H for whose $x_i \neq 0$; thus we have $Ham(x, 0)$ columns of H which are linearly dependent.

On the contrary assume that we are given $1 \leq i_1 < \cdots < i_t \leq n$ such that $\sum_{j=1}^{t} H_{\bullet i_j} x_{i_j} = 0$ for given $x_{i_j} \in \mathbb{F} \setminus \{0\}$, $j = 1, \ldots, t$, i. e. the columns $i_1, \ldots, i_t$ are linearly dependent. Let $x_i = 0$ if $i \neq i_j$ for all j; then $x \in \mathbb{F}^n$ and

$$\sum_{j=1}^{t} H_{\bullet i_j} x_{i_j} = \sum_{i=1}^{n} H_{\bullet i} x_i,$$

and thus $x = (x_i)_i \in C$ with $Ham(x, 0) = t$. □

Corollary 1.11.32 (Singleton Bound). *An $[n,k]$ linear code C has distance $d(C) \leq n-k+1$.*

Proof. The columns of any parity check matrix H of C are vectors in $\mathbb{F}^{n-k}$, and thus any selection of $n-k+1$ of them is linearly dependent. □

The *consequence* of this is that in order to have codes with good distances, it is desireable to construct matrices H such that only $(d-1)$ columns are linearly independent.

Example 1.11.33 (the Hamming code). *Let*

$$H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix} \in \mathbb{F}_2^{3\times 7} \qquad \text{and} \qquad C = \ker H \subseteq \mathbb{F}_2^7.$$

In fact, the columns of H are all non-zero vectors of $\mathbb{F}_2^3$. A simple calculation shows $\dim C = 7-3 = 4$, and thus $|C| = 2^4 = 16$. Moreover $d(C) = 3$ by the lemma. The singleton bound is $n-k+1 = 4$, and so one may ask whether this code is good or not.

The sphere packing bound computes to

$$128 = 16 \cdot \left(1 + \binom{7}{1} \cdot 1\right) = |C| \cdot |B_1(0)| \leq 2^7 = 128,$$

and thus the code is optimal! Such codes with attain the sphere packing bound are called perfect*.*

In order to reach the singleton bound it is neccessary to have larger field sizes. It has to be that $|\mathbb{F}| \geq n-2$.

Example 1.11.34. *Assume $\alpha_1, \ldots, \alpha_n$ are pairwise distinct non-zero elements of a finite field $\mathbb{F}$. Consider the matrix*

$$H = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \\ \alpha_1^2 & \alpha_2^2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \alpha_1^{n-k} & \alpha_2^{n-k} & \cdots & \alpha_n^{n-k} \end{pmatrix} \in \mathbb{F}^{(n-k)\times n}, \qquad \text{and } C = \ker H \subseteq \mathbb{F}^n.$$

By exploiting the Vandermonde determinant one easily shows that every $(n-k)$ columns of H are linearly independent, and thus $d(C) = n-k+1$—the singleton bound is attained! An example for this is the ISBN code.

1.11.4.2 The McEliece System

Some *facts* about codes:

1) Decoding a general $[n,k]$ linear code is an NP-hard problem.

2) There are special classes of codes going under the name of *algebraic geometric codes* (including Reed-Solomon and BCH codes) where decoding can be achieved in polynomial time.

The *idea* of McEliece was: take an algebraic geometric code C in $\mathbb{F}_2^n$ which can be efficiently decoded, and choose a generator matrix G of C. Let $T \in GL_k(\mathbb{F}_2)$ be a random invertible matrix, and $P \in GL_n(\mathbb{F}_2)$ a random permutation matrix. Compute $\tilde{G} := TGP$, and choose an t such that $d(C) \geq 2t+1$.

The *private* key consists of G, T, P and C, and the *public key* of $\tilde{G}$ and t.

For *encrypting* a plaintext $m \in \mathbb{F}_2^k$, choose a random $e \in \mathbb{F}_2^n$ such that $Ham(e,0) \leq t$, and let the ciphertext c be

$$c := c(m) := m\tilde{G} + e.$$

For decryption, apply the inverses of G and T to c to get $mG + \tilde{e}$, where $Ham(\tilde{e},0) \leq t$, and any decoding scheme for C can be applied.

Suggested sizes are $n = 1000$ and $k = 500$, resulting in a public key of $n \cdot k = 500000$ bits, which is approximately 61 kilobyte.

1.11.5 One-Way Trapdoor Functions from Semigroup Actions

Definition 1.11.35. *A* semigroup *is a set with an associative multiplication.*

Remark 1.11.36. *In general a semigroup has no identity element and thus there exist no inverses.*

Example 1.11.37. *For example, take the even integers with multiplication, i. e. the set* $2\mathbb{Z}$ *with multiplication* $\cdot$*.*

Definition 1.11.38. *Let* G *be a semigroup and* S *an arbitrary set. An* action *of* G *on* S *is a map* $\psi : G \times S \to S$ *such that*

$$\psi(a, \psi(b, s)) = \psi(ab, s) \qquad \text{for all } a, b \in G \text{ and } s \in S.$$

We will use the notation as instead of $\psi(a, s)$ *for* $a \in G$*,* $s \in S$*.*

Example 1.11.39. *Let* $G = 2\mathbb{Z}$ *be a semigroup with respect to the usual multiplication, and let* $S = E(\mathbb{F}_q)$ *be the set of points on an elliptic curve over* $\mathbb{F}_q$*. Then* $\psi : G \times S \to S$*,* $(y, P) \mapsto yP$ *defines a semigroup action of* G *on* S*.*

Application 1.11.40 (Extended Diffie-Hellman Key Exchange). *Alice and Bob agree on an* Abelian *semigroup* G *which acts on a set* S*, and they agree on an* $s \in S$*. Alica chooses an* $a \in G$ *and publishes* as*, and Bob chooses an* $b \in G$ *and publishes* bs*. The common key is* $k = (ab)s = a(bs) = b(as)$*.*

Note that this generalizes the usual Diffie-Hellman exchange: indeed, if H *is a group, then* $\psi : \mathbb{Z} \times H \to H$*,* $(n, h) \mapsto n^h$ *is a semigroup action, as* $(h^n)^m = h^{nm} = (h^m)^n$*.*

Application 1.11.41 (Extended ElGamal One-Way Trapdoor Function). *Given an Abelian semigroup action* $\psi : G \times S \to S$ *where* S *has the structure of a group with operation* $\circ$*. (This structure does not have to be compatible with* ψ*.)*

First Alice chooses an $s \in S$ *and* $a \in G$*, and publishes* (s, as)*. Her private key is* a*. To send a secret messge to Alice, Bob randomly chooses an element* $b \in G$ *and encrypts* $m \in S$ *by*

$$(m, b) \mapsto (bs, (b(as)) \circ m) = (c_1, c_2) \in S^2.$$

Alice can compute m *from* (c_1, c_2) *by*

$$m = (b(as))^{-1} \circ (b(as)) \circ m = (a(bs))^{-1} \circ c_2 = (ac_1)^{-1} \circ c_2.$$

Remark 1.11.42. *The difficulty of both extended Diffie-Hellman and ElGamal is based on the* semigroup action problem *(SAP) which asks for a semigroup action* $\psi : G \times S \to S$*:*

Given s *and* as*, find an* $\alpha \in G$ *such that* $\alpha s = as$*.*

For example in the case of Diffie-Hellman assume that Eve finds an $\alpha \in G$ *such that* $\alpha s = as$*. Then Eve can compute* $k = (ab)s = b(as) = b(\alpha s) = \alpha(bs)$*.*

Thus we have to require that the SAP is a hard problem. For this $\psi : G \times S \to S$ is a semigroup action as above. Define

$$G_{Eve} = \{\alpha \in G \mid \alpha s = as\}.$$

When G is a group, one has

$$\begin{aligned} \alpha \in G_{Eve} &\Leftrightarrow a^{-1}\alpha s = s \\ &\Leftrightarrow a^{-1}\alpha \in \mathrm{Stab}(s) \\ &\Leftrightarrow \alpha\,\mathrm{Stab}(s) = a\,\mathrm{Stab}(s) \\ &\Leftrightarrow \alpha \in a\,\mathrm{Stab}(s), \end{aligned}$$

where $\mathrm{Stab}(s) = \{g \in G \mid gs = s\}$ is the *stabilizer* of $s \in S$. (Note that $\mathrm{Stab}(s)$ is a groubgroup of G in this case.)

Lemma 1.11.43. *If G is a group, then* $\mathrm{Orbit}(s) = Gs$, *which is bijective to* $G/\mathrm{Stab}(s)$. *Here* $\mathrm{Orbit}(s) = \{gs \mid g \in G\}$ *is the* orbit *of $s \in S$ under G.*

When G is only a semigroup, one still has $a\,\mathrm{Stab}(s) \subseteq G_{Eve}$. In this case, if G is finite one has that $|Gs| \leq |G/\mathrm{Stab}(s)| = \frac{|G|}{|\mathrm{Stab}(s)|}$. In order to avoid brute force attacks we want $|Gs|$ to be a large. For example, $|Gs|$ should be $\geq 2^{80}$.

We want to give two examples of semigroup actions.

Examples 1.11.44.

(a) Linear Algebra Action*: Let $G = \mathbb{F}^{n\times n}$ where $\mathbb{F}$ is a field, and $S = \mathbb{F}^n$ with the action $\psi(A, v) = Av$. (Clearly we have $A(Bv) = (AB)v$.) The only problem is that for $n > 1$ we have that G is not Abelian.*

Given an $A \in \mathbb{F}^{n\times n}$, we want to find the largest commutative semi-subgroup of G such that $A \in G \subseteq \mathbb{F}^{n\times n}$. If A has finite order m, then $\{E_n, A, \ldots, A^{m-1}\}$ would be an Abelian subgroup. But is it the largest one?

Lemma 1.11.45. *If A is diagonizable with pairwise distinct eigenvalues $\lambda_1, \ldots, \lambda_n$, then the largest Abelian semi-subgroup containing A is*

$$\mathbb{F}[A] = \{f(A) \mid f \in \mathbb{F}[x]\}.$$

Outline of the Proof. Let A be as in the claim, and assume

$$SAS^{-1} = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{pmatrix}$$

for some invertible $S \in \mathbb{F}^{n\times n}$. First note that $\mathbb{F}[A]$ is an Abelian semi-subgroup containing A. Given B with $AB = BA$, we have

$$SAS^{-1}SBS^{-1} = SBS^{-1}SAS^{-1},$$

where with $\tilde{B} = SBS^{-1}$ we get that $\tilde{B}$ must be a diagonal matrix (why?). Thus $\tilde{B} \in \mathbb{F}[SAS^{-1}]$ and therefore $B = S^{-1}\tilde{B}S \in \mathbb{F}[A]$. □

An attempt would be to use such an $A \in \mathbb{F}^{n\times n}$ and $G = \mathbb{F}[A]$ as an Abelian semigroup acting on $S = \mathbb{F}^n$. But then Eve can solve the SAP using linear algebra:

Recall the Theorem of Cayley-Hamilton: if $\chi(A) = \det(xE_n - A) \in \mathbb{F}[x]$ is the characteristic polynomial of A, then $\chi(A) = 0$. (Hence the minimal polynomial of A exists and divides χ.)

Given $A \in \mathbb{C}^{n\times n}$ with eigenvalues $\lambda_1, \ldots, \lambda_m \in \mathbb{C}$ and multiplicities $\mu_1, \ldots, \mu_m \in \mathbb{N}$, and two analytic (convergent) functions $f, g \in \mathbb{C}[x]$ such that $f^{(j)}(\lambda_i) = g^{(j)}(\lambda_i)$ for all $1 \leq i \leq m$ and $0 \leq j < \mu_i$, then $f(A) = g(A)$. (For $\mu_i = 1$, $1 \leq i \leq n$, this directly follows from $Sf(A)S^{-1} = f(SAS^{-1})$ by using an $S \in \mathbb{C}^{n\times n}$ such that SAS^{-1} is a diagonal matrix.)

If Eve knows the semigroup action $\mathbb{F}[A] \times \mathbb{F}^n \to \mathbb{F}^n$ and $v \in \mathbb{F}^n$, $Bv \in \mathbb{F}^n$ for some $B \in \mathbb{F}[A]$. By Cayley-Hamilton she knows $B = \sum_{i=0}^{n-1} a_i A^i$ (if $B = f(A)$ for some $f \in \mathbb{F}[x]$, write $f = \chi_A \cdot g + h$ for $g, h \in \mathbb{F}[x]$, $\deg h < \deg \chi_A = n$, then $B = f(A) = h(A)$), and thus $Bv = \sum_{i=0}^{n-1} a_i(A^i v)$. But then Eve can simply compute v from Bv and B (and therefore from the a_i's) by use of ordinary linear algebra!

(b) *A more interesting example is the following: Take $S = E(\mathbb{F}_q) \times E(\mathbb{F}_q)$ for some elliptic curve E over a finite field $\mathbb{F}_q$. Let $G = \mathbb{Z}[A]$, where $A \in \mathbb{Z}^{2\times 2}$. The action is given by*

$$G \times S \to S, \qquad \left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, (P, Q)\right) \mapsto (aP + bQ, cP + dQ).$$

Remark 1.11.46. *If P, Q are in the same orbit, i. e. $Q = nP$, then the above problem can be solved by several DLPs in $E(\mathbb{F}_q)$. In the general case we do not know how to reduce this problem to DLPs in $E(\mathbb{F}_q)$.*

(c) *Back to the linear algebra problem. This problem becomes more complicated if one works with a finite ring and not a finite field. Moreover note that a matrix multiplication is defined such that it is enough to have a finite* semiring $(R, +, \cdot)$, *i. e. a structure such that*

(i) $(R, +)$ *is an Abelian semigroup,*

(ii) $(R, \cdot)$ *is a semigroup,*

(iii) *the usual distributive laws hold.*

A congruence relation *on a semiring R is an equivalence relation $\sim$ on R satisfying that if $a \sim b$ holds for any $a, b, c \in R$, we also have $ac \sim bc$, $ca \sim cb$ and $a + c \sim b + c$. Clearly $\{(a, a) \mid a \in R\}$ and R^2 are congruence relations on R for every semiring R; they are called the* trivial *ones. A semiring without nontrivial congruence relations is called* simple.

For example, the ring $R = \mathbb{Z}/6\mathbb{Z}$ is not simple, since it possesses the non-trivial congruence relations induced by the ideals $2R$ and $3R$.

Example 1.11.47. *A simple semiring with six elements:*

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

$\cdot$	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	2	0	0	5
3	0	3	4	3	4	3
4	0	4	4	0	0	3
5	0	5	2	5	2	5

1.12 Lattices and the LLL Algorithm

Given vectors $v_1, \dots, v_k \in \mathbb{R}^n$ which are linearly independent, one defines:

Definition 1.12.1. *The set*

$$\Lambda = \Lambda(v_1, \dots, v_k) := \sum_{i=1}^{k} \mathbb{Z} v_i = \left\{ \sum_{i=1}^{k} \lambda_i v_i \mid \lambda_1, \dots, \lambda_k \in \mathbb{Z} \right\}$$

is called a (k-dimensional) lattice.

Remark 1.12.2. *Obviously* Λ *is a* $\mathbb{Z}$*-module. Moreover, it is free of rank* k*, i. e. it is isomorphic to* $\mathbb{Z}^k$*.*

Assume $v_i = (v_{i1}, \dots, v_{in})$*; then* Λ *is the row space* $\operatorname{rowsp}_{\mathbb{Z}} M$ *of the matrix*

$$M := \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix},$$

and $M = M(v_1, \dots, v_n)$ *is called a* generator matrix *of* $\Lambda = \Lambda(v_1, \dots, v_n)$*.*

Lemma 1.12.3. *Let* M *and* $\tilde{M}$ *be two generator matrices. Then* M *and* $\tilde{M}$ *generate the same lattice if and only if there exists a unimodular*[13] $U \in GL_k(\mathbb{Z})$ *satisfying* $\tilde{M} = UM$*.*

Proof. If $\tilde{M} = UM$ with an $U \in GL_k(\mathbb{Z})$, one easily sees that $\Lambda(M) = \Lambda(\tilde{M})$ since $v \mapsto Uv$ is an isomorphism of $\mathbb{Z}^k$.

Conversely assume that $\operatorname{rowsp}_{\mathbb{Z}} M = \operatorname{rowsp}_{\mathbb{Z}} \tilde{M}$, and let $M = M(v_1, \dots, v_k)$ and $\tilde{M} = M(\tilde{v}_1, \dots, \tilde{v}_k)$. Write $\tilde{v}_j = \sum_{i=1}^{k} \lambda_{i,j} v_i$, where $\lambda_{i,j} \in \mathbb{Z}$; then $e_i \mapsto \sum_{j=1}^{n} \lambda_{i,j} e_j$ defines a map from $\mathbb{Z}^k \to \mathbb{Z}^k$ which easily can be seen to be an isomorphism. Thus $U = (\lambda_{i,j})_{i,j} \in \mathbb{Z}^{k \times k}$ is unimodular and $\tilde{M} = UM$. □

Given a lattice $\Lambda = \operatorname{rowsp}_{\mathbb{Z}}(M)$, one defines the *volume of the fundamental region* as

$$\left|\det(MM^t)\right|^{1/2}.$$

This expression is sometimes also called the *determinant* of the lattice. Note that in the literature sometimes $\left|\det(MM^t)\right|$ is defined to be the determinant.

Remark 1.12.4. *If* $\tilde{M} = UM$ *with* $U \in GL_2(\mathbb{Z})$*, then* $\det(\tilde{M}\tilde{M}^t) = \det U \cdot \deg(MM^t) \cdot \det U^t = \det(MM^t)$ *since* $\det U \cdot \det U^t = (\det U)^2 = 1$*.*

Definition 1.12.5. *Let* Λ *be a lattice. Then define*

$$d := d_\Lambda := \min\{\|x - y\| \mid x, y \in \Lambda,\ x \neq y\} \qquad \textit{and} \qquad r := r_\Lambda := \frac{d_\Lambda}{2}.$$

Then one defines the density of Λ *as*

$$\operatorname{density}(\Lambda) := \frac{\text{vol}(k\textit{-ball of radius } r)}{\sqrt{\det(MM^t)}}.$$

Example 1.12.6 (Convay and Sloane, "Codes and Lattices"). *Let* Λ *be the lattice defined by*

$$M := \begin{pmatrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2}\sqrt{3} \end{pmatrix}.$$

(See figure 1.8.) Then

$$\det MM^t = \det \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix} = \frac{3}{4}, \qquad \textit{and hence} \qquad (\det MM^t)^{1/2} = \frac{\sqrt{3}}{2} = |\det M|.$$

We get

$$d = 1, \quad r = \frac{1}{2} \qquad \textit{and thus} \qquad \operatorname{density}(\Lambda) = \frac{\pi(1/2)^2}{\sqrt{3}/2} = \frac{\pi}{2\sqrt{3}} \approx 0.906.$$

[13] A matrix $M \in R^{n \times n}$ over a ring R is called *unimodular* if it is invertible, i. e. if $M \in GL_n(R)$. By Cramer's Rule one easily shows that this is the case if and only if $\det M \in R^*$.

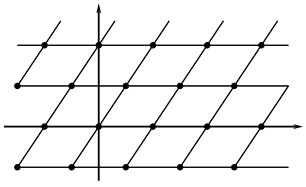


Figure 1.8: The lattice generated by the vectors $(1,0)$ and $(1/2, \sqrt{3}/2)$

A big *question* in lattice theory is:

Question 1.12.7. *Given a lattice Λ, find the smallest non-zero lattice vector v, i. e. a vector $v \in \Lambda$ such that $\|v\| = \min\{\|w\| \mid w \in \Lambda,\ w \neq 0\}$.*

This is equivalent to computing a vector in Λ of length d_Λ, and it is known to be a NP-hard problem for the ∞-norm; in the case of the Euclidean 2-norm it still unknown.

In the case of n-dimensional lattices this means M is a square matrix and $|\det M| = \left|\det(MM^t)\right|^{1/2}$. One has the following estimate of the smallest vector due to Hermite: it is

$$d \leq c \cdot (\det M)^{1/4}, \qquad \text{where } c = \frac{1}{\sqrt{\pi e}} \text{ is the } \textit{Hermite constant.}$$

Facts about n-dimensional lattices

- Finding the shortest vector $v \in \Lambda$ is a very hard problem in general.
- In 1982, Lenstra, Lenstra and Lovacs derived an algorithm called LLL or L^3 which transforms M to a so called Lovacs basis. (We will see later what that is.) The smallest basis vector b has norm
$$\|b\| \leq 2^{n/2} \left|\det M\right|^{1/n}.$$
In fact, if the density of the lattice "is low", it can be shown that LLL even finds a shortest vector. For an application of this fact, see the Lagarias-Odlyzko attack on the Merkle-Hellman cryptosystem in section 1.11.2.

Norms

Definition 1.12.8. *For every positive $p \geq 1$ and $w = (w_1, \ldots, w_n) \in \mathbb{R}^n$, define the p-*norm

$$\|w\|_p := \left(\sum_{i=1}^{n} |w_i|^p\right)^{1/p}.$$

Moreover let

$$\|w\|_\infty := \max_{i=1,\ldots,n} |w_i|$$

*be the ∞-*norm *of w.*

Remark 1.12.9. *For every $p \in [1, \infty]$ we have that $\|\cdot\|_p$ is a norm on $\mathbb{R}^n$, i. e. the following holds:*

(i) *we have $\|w\|_p \geq 0$ for all $w \in \mathbb{R}^n$, and $\|w\|_p = 0$ if and only if $w = 0$.*

(ii) *we have $\|v + w\|_p \leq \|v\|_p + \|w\|_p$ for all $v, w \in \mathbb{R}^n$,*

(iii) *we have $\|\lambda w\|_p = |\lambda| \, \|w\|_p$ for all $\lambda \in \mathbb{R}$ and $w \in \mathbb{R}^n$.*

Remark 1.12.10. *Different norms on $\mathbb{R}^n$ induce the same topologies. For example,*

$$\|x\|_\infty \leq \|x\|_2 \leq \|x\|_1 \leq \sqrt{n}\, \|x\|_2 \leq n\, \|x\|_\infty$$

for every $x \in \mathbb{R}^n$.

Relation of Lattices to Quadratic Forms

Definition 1.12.11. *An expression of the form*

$$q(x) = \sum_{1 \le i \le j \le n} q_{ij} x_i x_j \in \mathbb{R}[x_1, \ldots, x_n]$$

is called an quadratic form. *Introduce a matrix*

$$S = \begin{pmatrix} q_{11} & & \frac{q_{ij}}{2} \\ & \ddots & \\ \frac{q_{ji}}{2} & & q_{nn} \end{pmatrix} \in \mathbb{R}^{n \times n}.$$

Then $q(x) = xSx^t$ for every $x \in \mathbb{R}^n$. A quadratic form q is positive definite *if $q(x) \ge 0$ for every $x \in \mathbb{R}^n$, and if $q(x) = 0$ if and only if $x = 0$. Furthermore two quadratic forms q and $\tilde{q}$ are* equivalent *if there is an orthogonal matrix $U \in \mathbb{R}^{n \times n}$ such that $\tilde{q}(x) = q(Ux)$ for every $x \in \mathbb{R}^n$.*

Remark 1.12.12. *Since $q(xU) = xUSU^t x^t = x(USU^t)x^t$, transformation by U corresponds to $S \to USU^t$.*

Consider a lattice $\Lambda = \mathrm{rowsp}_{\mathbb{Z}}(M)$, and define $S := MM^t$. This defines a positive definite quadratic form $q_\Lambda(x) := xMM^t x^t$, $x \in \mathbb{R}^n$, and two such forms are q, $\tilde{q}$ belong to the same lattice if and only if they are *congruent*.

A study of lattices is hence closely related to the study of quadratic forms.

Fundamental Problems Associated to Lattices

1. The *Shortest Vector Problem (SVP)*: Given $\Lambda = \Lambda(v_1, \ldots, v_k)$ and $p \in \mathbb{N}_{>0} \cup \{\infty\}$, find a "shortest vector $x \in \Lambda$" with respect to the p-norm. This is a vector $x \ne 0$ such that

   $$\|x\|_p = \inf_{\substack{y \ne 0 \\ y \in \Lambda}} \|y\|_p .$$

   Note that for $p = \infty$ this problem is known to be NP-complete. For $p = 2$ the complexity is unknown, but recent results by Miklos Ajtai indicate that this problem is hard.

2. The *Closest Vector Problem (CVP):* Given a lattice Λ and $p \in \mathbb{N}_{>0} \cup \{\infty\}$, and a $w \in \mathbb{R}^n$, find an $x \in \Lambda$ such that

   $$\|x - w\|_p = \inf_{y \in \Lambda} \|y - w\|_p .$$

   Even for $p = 2$, the CVP is known to be NP-complete.

An application of the SVP is the knapsack problem (note Section 1.11.2). Recall that for this problem one is given positive numbers $a_1, \ldots, a_n \in \mathbb{N}$ and $s \in \mathbb{N}$, and one is asked whether there is a subset $S \subseteq \{1, \ldots, n\}$ such that

$$\sum_{i \in S} a_i = s.$$

For different choices of N consider the lattice generated by

$$M = \left(\begin{array}{ccc|c} & & & -Na_1 \\ & E_n & & \vdots \\ & & & -Na_n \\ \hline 0 & \cdots & 0 & Ns \end{array}\right) \in \mathbb{R}^{(n+1)\times(n+1)}.$$

If the knapsack problem has a solution, then it follows that the lattice $\mathrm{rowsp}_{\mathbb{Z}}(M)$ has a shortest vector x of ∞-norm 1 and such that $0 \le x_i$ for all components x_i of x. (This relation goes back to Lagarias and Odlyzko; later C. Schnorr showed how to get a factorization method from this.)

Hermite Basis Given a $k \times n$-matrix $M = (a_{ij})_{ij}$ of rank m, one says M is in *Hermite form* if there are numbers $1 \leq j_1 < \cdots < j_m \leq n$ such that

a) we have $a_{ij} = 0$ for all $i > m$, or if $i \leq m$ for all $j < j_i$;

b) we have $a_{ij_i} > 0$ for all $i \leq m$;

c) we have $0 \leq a_{kj_i} < a_{ij_i}$ for all $k < i \leq m$.

Thus M looks like

$$M = \begin{pmatrix} \boxed{a_{1j_1}} & * & < & * & < & * \\ & & \boxed{a_{2j_2}} & * & \vdots & * \\ & & & \ddots & < & \\ & & & & \boxed{a_{mj_m}} & * \\ 0 & & & & & \end{pmatrix}.$$

Theorem 1.12.13 (Hermite). *Let $M \in \mathbb{Q}^{k\times n}$ be a matrix of rank m. Then there is a unimodular matrix $U \in GL_k(\mathbb{Z})$ such that $\tilde{M} = UM$ is in Hermite form. Moreover the form is unique.*

Outline of Proof. First note that by multiplying with the least common multiple of all denominators we reduce to the problem that $M \in \mathbb{Z}^{k\times n}$.

Let j_1 be the first column with nonzero entries. Using only unimodular row operations and Euclids algorithm for the greatest common divisor it is possible to transform M into

$$\tilde{M} = \begin{pmatrix} 0 & g & M_1 \\ 0 & 0 & M_2 \end{pmatrix},$$

where $g > 0$ is the greatest common divisor of all entries in the j_1-th column. By recursion on M_2 we get a form

$$\begin{pmatrix} a_1 & * & * & * & * & \\ & & a_2 & * & * & \\ & & & & a_3 & \\ & & & & & \ddots \end{pmatrix}$$

where a_i is the greatest common divisor of all entries in the j_i-th column. By subtracting rows one can transform the $*$'s in the j_i-th columns to integers in the interval $[0, a_i[$.

The uniqueness part is left as an exercise to the reader. □

Remark 1.12.14. *Consider a special case where $m = k = n$ (thus we have an $n \times n$ invertible matrix with rational entries). Then Hermite says that there is a $T \in GL_n(\mathbb{Z})$ such that*

$$\tilde{M} = TM = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ & \ddots & \\ 0 & & a_{nn} \end{pmatrix}$$

with $a_{ii} > 0$ for all i and $0 \leq a_{ij} < a_{jj}$ for all $1 \leq i < j \leq n$. Especially the theorem provides a unique basis for row modules. Also note that in general the norm of the basis elements are "fairly small".

Remark 1.12.15. *The Hermite Theorem is not true if some entries of M are irrational; for example consider the matrix*

$$\begin{pmatrix} 1 & 3 \\ \sqrt{2} & 4 \end{pmatrix} \in \mathbb{R}^{2\times 2}.$$

Question 1.12.16. *How to get a smaller basis of* $\mathrm{rowsp}_{\mathbb{Z}}(M)$ *than the one provided by Hermite?*

The answer came 1982 and was by Lenstra, Lenstra and Lovacs: the L^3 or LLL algorithm. The idea is based on the Gram-Schmidt algorithm for orthogonalization, which we will explain first.

Gram-Schmidt Orthogonalization Let $v_1, \ldots, v_n \in \mathbb{R}^n$ be a basis of $\mathbb{R}^n$. Inductively define

$$v_i^* := v_i - \sum_{j=1}^{i-1} \mu_{ij} v_j^*, \qquad \text{where } \mu_{ij} = \frac{\left\langle v_i, v_j^* \right\rangle}{\left\langle v_j^*, v_j^* \right\rangle}.$$

Then we have:

(a) The vectors $v_1^*, \ldots, v_n^*$ form an orthogonal basis of $\mathbb{R}^n$;

(b) We have $U_k := \operatorname{span}\{v_1, \ldots, v_k\} = \operatorname{span}\{v_1^*, \ldots, v_k^*\}$ for all $1 \leq k \leq n$;

(c) We have that v_i^* is the projection of v_i onto

$$U_i^{\perp} = \{x \in \mathbb{R}^n \mid \langle x, v_j \rangle = 0 \text{ for } j = 1, \ldots, n\}.$$

In particular we have $\|n_i^*\|_2 \leq \|v_i\|_2$.

Proof. Clearly (a) follows from (b). For (2) note that by definition we have $\operatorname{span}\{v_1^*, \ldots, v_k^*\} \subseteq U_k$ for all k. Since $v_i \notin \operatorname{span}\{v_1, \ldots, v_{i-1}\}$ clearly $v_i^* \neq 0$. Now $\left\langle v_i^*, v_j^* \right\rangle = 0$ (see below) for $i \neq j$, and therefore the v_i^* are linearly independent and the claim follows by induction.

To see that $\langle v_k^*, v_i^* \rangle = 0$, we proceed by induction first on k and then on i, where $1 \leq k < i \leq n$. We get

$$\langle v_k^*, v_i^* \rangle = \left\langle v_k^*, v_i - \sum_{j=1}^{i-1} \mu_{ij} v_j^* \right\rangle = \langle v_k^*, v_i - \mu_{ik} v_k^* \rangle = \langle v_k^*, v_i \rangle - \frac{\langle v_k^*, v_i \rangle}{\langle v_k^*, v_k^* \rangle} \langle v_k^*, v_k^* \rangle = 0.$$

For (c) note that since $v_i = v_i^* + (v_i - v_i^*)$ and $v_i^* \in U_{i-1}^{\perp}$ we have to show that $v_i - v_i^* \in U_{i-1}$. But

$$v_i - v_i^* = \sum_{j=1}^{i-1} \mu_{ij} v_j^*.$$

Moreover

$$\|v_i\|_2^2 = \|v_i^*\|_2^2 + \|v_i - v_i^*\|_2^2 \geq \|v_i^*\|_2^2,$$

since v_i^* and $v_i - v_i^*$ are orthogonal. □

Lemma 1.12.17. *Let*

$$B = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} = \begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{n1} & \cdots & v_{nn} \end{pmatrix}$$

and

$$B^* = \begin{pmatrix} v_1^* \\ \vdots \\ v_n^* \end{pmatrix} = \begin{pmatrix} v_{11}^* & \cdots & v_{1n}^* \\ \vdots & \ddots & \vdots \\ v_{n1}^* & \cdots & v_{nn}^* \end{pmatrix}.$$

Then

$$B = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \mu_{21} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \mu_{n1} & \cdots & \mu_{n,n-1} & 1 \end{pmatrix} B^*.$$

In particular $\det B = \det B^* \neq 0$.

Proof. Clear. □

Theorem 1.12.18 (Hadamard). *Let* $B = \left(v_1, \ldots, v_n\right)^t = (v_{ij})_{ij} \in \mathbb{R}^{n \times n}$ *be as above. Let* $\rho \in \mathbb{R}_{>0}$ *be such that* $|v_{ij}| \leq \rho$ *for all* i, j. *Then*

$$|\det B| \leq \|v_1\|_2 \cdots \|v_n\|_2 \leq n^{n/2} \rho^n.$$

Proof. For the first inequality note that

$$|\det B| = |\det B^*| = \|v_1^*\|_2 \cdots \|v_n^*\|_2 \leq \|v_1\|_2 \cdots \|v_n\|_2 ,$$

since B^* is orthogonal:

$$|\det B^*| = \sqrt{|\det B^*(B^*)^t|} = \sqrt{\deg \begin{pmatrix} \|v_1\|_2^2 & & 0 \\ & \ddots & \\ 0 & & \|v_n\|_2^2 \end{pmatrix}}.$$

For the second inequality note that $\|v_i\|_2 \leq \sqrt{n}\,\|v_i\|_\infty \leq \sqrt{n}\rho$. □

Definition 1.12.19. *A matrix B having entries v_{ij} with $|v_{ij}| \leq \rho$ is called a* Hadamard matrix *if equality holds, i. e. if $|\det B| = n^{n/2}\rho^n$.*

The following is a direct consequence of the previous lemmas: if equality holds, then we have

(1) all entries v_{ij} satisfy $|v_{ij}| = \rho$,

(2) the rows are pairwise orthogonal.

Examples 1.12.20.

1. *For $n = 2$ and $\rho = 1$ we can take*

$$H_2 = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

2. *For $n = 3$ and $\rho = 1$ there is no such matrix.*

3. *For $n = 4$ and $\rho = 1$ we can take*

$$\begin{pmatrix} H_2 & -H_2 \\ H_2 & H_2 \end{pmatrix} = \begin{pmatrix} 1 & -1 & -1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 \end{pmatrix}.$$

4. *For $n = 2^k$, $k > 0$ and $\rho = 1$ we can define*

$$H_{2^k} = \begin{pmatrix} H_{2^{k-1}} & -H_{2^{k-1}} \\ H_{2^{k-1}} & H_{2^{k-1}} \end{pmatrix} \in \mathbb{R}^{2^k \times 2^k}.$$

Lemma 1.12.21. *Let $\Lambda = \Lambda(v_1, \ldots, v_n) = \sum_{i=1}^n \mathbb{Z} v_i$ be a n-dimensional lattice, and let $v_1^*, \ldots, v_n^* \in \mathbb{R}^n$ be the associated Gram-Schmidt basis. (Note that $v_i^* \notin \Lambda$ in general.) If $v \in \Lambda$ is any nonzero vector, then*

$$\|v\|_2 \geq \min\{\|v_1^*\|_2, \ldots, \|v_n^*\|_2\}.$$

Proof. Let $v = \sum_{i=1}^n c_i v_i$ for $c_i \in \mathbb{Z}$, and assume ℓ is the highest index such that $c_\ell \neq 0$. Then

$$v = \sum_{i=1}^{\ell} c_i v_i = \sum_{i=1}^{\ell} c_i \left(v_i^* + \sum_{j=1}^{i-1} \mu_{ij} v_j^* \right) = c_\ell v_\ell^* + \sum_{j=1}^{\ell-1} r_j v_j^*$$

for appropiate $r_j \in \mathbb{R}$. Using the orthogonality of the v_i^*'s we get

$$\|v\|_2^2 = |c_\ell|^2 \|v_\ell^*\|_2^2 + \sum_{j=1}^{\ell-1} |r_j|^2 \left\|v_j^*\right\|_2^2 \geq \|v_\ell^*\|_2^2 \geq \min\{\|v_1^*\|_2, \ldots, \|v_n^*\|_2\}.$$

□

As a consequence, $\min\{\|v_1^*\|_2, \ldots, \|v_n^*\|_2\}$ gives a lower bound for the shortest vector problem. Moreover it motivates the idea that an "approximate Gram-Schmidt" (such that the transformed basis is still in the lattice) should give a "short basis".

Definition 1.12.22. *The ordered basis $(v_1, \ldots, v_n) \in \mathbb{R}^{n\times n}$ is called* length reduced *if $|\mu_{ij}| \leq \frac{1}{2}$ for $i \neq j$ when doing Gram-Schmidt.*

Theorem 1.12.23. *If $(v_1, \ldots, v_n) \in \mathbb{R}^{n\times n}$ is length reduced, then*

$$\|v_i\|_2^2 \leq \|v_i^*\|_2^2 + \frac{1}{4}\sum_{j=1}^{i-1} \|v_j^*\|_2^2, \qquad i = 1, \ldots, n.$$

Proof. By definition we have that

$$v_i = v_i^* + \sum_{j=1}^{i-1} \mu_{ij} v_j^*,$$

and because of orthogonality we have

$$\|v_i^*\|_2^2 = \|v_i^*\|_2^2 + \sum_{j=1}^{i-1} |\mu_{ij}|^2 \|v_j^*\|_2^2 \leq \|v_i^*\|_2^2 + \frac{1}{4}\sum_{j=1}^{i-1} \|v_j^*\|_2^2.$$

□

Lemma 1.12.24. *Every lattice $\Lambda(v_1, \ldots, v_n)$ has a length reduced basis $\tilde{v}_1, \ldots, \tilde{v}_n$.*

Sketch of Proof. For $i = 2, 3, \ldots, n$ let

$$\tilde{v}_i := v_i - \sum_{j=1}^{i-1} \lfloor \mu_{ij} \rceil \, v_j,$$

where $\lfloor \alpha \rceil := \lceil \alpha + \frac{1}{2} \rceil$. □

Definition 1.12.25 (Lenstra, Lenstra, Lovacs 1982). *Given a lattice $\Lambda(v_1, \ldots, v_n)$ and a Gram-Schmidt basis $v_1^*, \ldots, v_n^*$, we say that $v_1, \ldots, v_n$ are* LLL-reduced with parameter ρ, *where $\frac{1}{4} < \rho < 1$, if*

(1) $|\mu_{ij}| \leq \frac{1}{2}$ *for $1 \leq i < j \leq n$ (i. e. the basis is length reduced) and*

(2) $\rho \left\|v_{k-1}^*\right\|_2^2 \leq \|v_k^*\|_2^2 + |\mu_{k,k-1}|^2 \left\|v_{k-1}^*\right\|^2$.

Theorem 1.12.26. *Let $v_1, \ldots, v_n$ be an LLL-reduced basis of a lattice Λ with parameter ρ. Define $\alpha = \frac{1}{\rho - 1/4}$. Then we have that*

(1) $\|v_1\|_2 \leq \alpha^{(n-1)/4} (\det \Lambda)^{1/n}$, *and*

(2) $\prod_{i=1}^n \|v_i\|_2^2 \leq \alpha^{\binom{n}{2}} (\det \Lambda)^2$.

(In the original paper from 1982, ρ was fixed to $3/4$, and thus $\alpha = 2$.)

Note that $(\det \Lambda)^{1/2}$ corresponds to the geometric mean of $\|v_1^*\|_2, \ldots, \|v_n^*\|_2$.

The LLL Algorithm Let $v_1, \ldots, v_n$ be a basis of the lattice $\Lambda = \Lambda(v_1, \ldots, v_n)$, and let $\rho \in \mathbb{R}$ such that $\frac{1}{4} < \rho < 1$. Then the LLL algorithm works as follows:

1. Let $k := 2$ and compute the Gram-Schmidt basis $v_1^*, \ldots, v_n^*$ and the μ_{ij}'s.

2. While $k \leq n$ do:

   (Assume that $v_1, \ldots, v_{k-1}$ is already LLL-reduced with parameter ρ.)

   (a) Length reduce $v_1, \ldots, v_{k-1}, v_k$ and recompute (if necessary) the μ_{ij}.

   (b) If $\rho \left\|v_{k-1}^*\right\|_2^2 > \|v_k^*\|_2^2 + |\mu_{k,k-1}|^2 \left\|v_{k-1}^*\right\|_2^2$ interchange v_k and v_{k-1} and set $k := \max\{k-1, 2\}$; otherwise let $k := k+1$.

Theorem 1.12.27. *Let $\Lambda = \Lambda(v_1, \ldots, v_n)$ be an* integer lattice, *i. e. we have $v_i \in \mathbb{Z}^n$. Then the LLL algorithm correctly computes an LLL-reduced basis of Λ. If for some $B > 0$ we have $\|v_i\|_2 \leq B$ for all i, then the number of arithmetic operations in $\mathbb{Q}$ is $\mathcal{O}(n^4 \log B)$.*

Outline of Proof. Let $D_j := (\det \Lambda(v_1, \ldots, v_j))^2$, $j = 1, \ldots, n$. Thus if $M_j = \left(v_1, \ldots, v_j\right)^t$ we have $D_j = \det(MM^t)$. Let $D = \prod_{j=1}^n D_j$; since Λ is an integer lattice we have $D \in \mathbb{N}$. During the process of length reduction we have $\operatorname{span}\{v_1, \ldots, v_i\} = \operatorname{span}\{\tilde{v}_1, \ldots, \tilde{v}_i\}$ and $\det \Lambda(v_1, \ldots, v_i) = \det \Lambda(\tilde{v}_1, \ldots, \tilde{v}_i\}$. Hence the length reduction does not changes D.

We claim that after the interchange in step 2(b) we have that $D_{new} \leq \frac{1}{\rho} D_{old}$:

It is

$$\rho \left\|v_{k-1,old}^*\right\|_2^2 > \left\|v_{k,old}^*\right\|_2^2 + |\mu_{k,k-1,old}|^2 \left\|v_{k-1,old}^*\right\|_2^2,$$

and therefore $\rho \left\|v_{k,new}^*\right\|_2 > \left\|v_{k-1,new}^*\right\|_2$. Note that $D_{k-1,old} = \left\|v_{1,old}^*\right\|_2^2 \cdots \left\|v_{k-1,old}^*\right\|_2^2$ and thus

$$D_{k-1,new} \leq \rho D_{k-1,old}.$$

Since D_k does not change, $D_{new} \leq \rho D_{old}$.

Initially the value of Di s bounded by

$$|D| = \left|\prod_{i=1}^n D_i\right| \leq \prod_{i=1}^n \left|B^2\right|^i = B^{n(n+1)};$$

hence the number of interchanges can be at most $\log_{1/\rho} B^{n(n+1)}$ or $\mathcal{O}(n^2 \log B)$. Recomputing the μ_{jk}'s needs $\mathcal{O}(k) \leq \mathcal{O}(n)$ for fixed k, and the length reduction requires $\mathcal{O}(n^2)$ arithmetic operations. Using this we can conclude. □

Remarks 1.12.28.

(1) *Finer analysis shows that the cost is $\mathcal{O}(n^4 \log B)$ arithmetic operations of integers having size at most $\mathcal{O}(n \log B)$.*

(2) *The theorem was concerned with integer lattices $\Lambda(v_1, \ldots, v_n) \subseteq \mathbb{Z}^n$. For rational lattices $\Lambda(v_1, \ldots, v_n) \subseteq \mathbb{Q}^n$ the same theorem holds after multiplying Λ by the least common multiple of the denominators of the components of all v_i's.*

(3) *The running time crucially depends on ρ. If $\rho = 1$ one can not use the above argument to show that the algorithm terminates in polynomial time. Experience shows that the algorithm converges to an LLL-reduced basis even for $\rho = 1$. Note that ρ close to 1 gives much better bases in general, and that the original paper (1982) only considered the special case $\rho = 3/4$.*

(4) *The important property of an LLL-reduced basis is that*

$$\|v_1\|_2 \leq \alpha^{(n-1)/4} (\det \Lambda)^{1/n},$$

where $\alpha = \frac{1}{\rho - 1/4}$.

1.13 Factoring

In the sequel let n be a composite positive integer. The goal is to find factors of n. If n has "small" factors, trial and error will "quickly" find them. Thus the hardest situtation seems to be $n = pq$, where p and q are distinct primes and have similar size.

Here trial and error is of exponential time, requiring $\mathcal{O}(\sqrt{n}) = \mathcal{O}(e^{\frac{1}{2}\log n})$ trials.

1.13.1 The Quadratic Sieve

The *basic idea* is to consider the currve $\{(\alpha, \beta) \in \mathbb{Z}_n^2 \mid \alpha^2 - \beta^2 = 0\}$. If one finds a nontrivial solution α, β, i. e. $\alpha \neq \pm\beta$, then we have

$$0 \equiv (\alpha + \beta)(\alpha - \beta) \pmod{n}$$

but

$$\alpha + \beta \not\equiv 0 \not\equiv \alpha - \beta \pmod{n}.$$

Thus either p divides $\alpha + \beta$ and q divides $\alpha - \beta$, or q divides $\alpha + \beta$ and p divides $\alpha - \beta$, and in any case

$$\{\gcd(\alpha + \beta, n), \gcd(\alpha - \beta, n)\} = \{p, q\}.$$

A *first approach* to find nontrivial solutions is to randomly search for $x_i \in \mathbb{Z}_n$ and hope that $x_i^2 \mod n = y^2$ for some $y \in \mathbb{Z}$ such that $y \not\equiv \pm x_i \pmod{n}$. The problem is that the chance that a random number $x_i^2 \mod n$ is a square in $\mathbb{Z}$ is approximately $\frac{1}{\sqrt{n}}$; thus this method has again exponential complexity.

An *improvement* is to let $p_1, \ldots, p_t$ be the first t primes, and to search for numbers x_i such that $x_i^2 \mod n = \prod_{j=1}^t p_j^{e_{ij}}$ for $e_{ij} \in \mathbb{N}$, i. e. such that $x_i^2 \mod n$ is p_t-smooth.

Note that if x_i^2 and x_j^2 can be factored, then we have

$$(x_i x_j)^2 \equiv \prod_{k=1}^{t} p_k^{e_{ik}+e_{jk}} \pmod{n}.$$

Moreover note that $x_i^2 \mod n$ is a square in $\mathbb{Z}$ if and only if e_{ij} is even for $j = 1, \ldots, t$.

Now assume that $x_1^2 \mod n, \ldots, x_k^2 \mod n$ are not squares. Form a matrix

$$A = \begin{pmatrix} \overline{e}_{11} & \cdots & \overline{e}_{1t} \\ \vdots & \ddots & \vdots \\ \overline{e}_{k1} & \cdots & \overline{e}_{kt} \end{pmatrix} \in \mathbb{F}_2^{k \times t},$$

where $\overline{\bullet} : \mathbb{Z} \to \mathbb{Z}/2\mathbb{Z} = \mathbb{F}_2$ is reduction modulo 2. If $k \geq t$ there is a good chance that A has a non-trivial left kernel. To every non-zero element of the left-kernel corresponds a number x where $x^2 \mod n$ is a square in $\mathbb{Z}$. In two third of the cases (which are $\alpha = \pm\beta$, p divides $\alpha + \beta$ or p divides $\alpha - \beta$) this leads to a solution. Thus the difficulty of factoring is reduced to find numbers $x \in \mathbb{Z}_n$ such that $x^2 \mod n$ is p_t-smooth.

Recall Theorem 1.9.11, which we will recite here:

Theorem 1.13.1 (Norton (1971), Canfield, Erdős, Pomerance (1983)). *Let N and r be positive reals satisfying*

$$B := N^{1/r} \geq \log N.$$

Then the number of $x \in \mathbb{N}$, $x \leq N$ which are B-smooth is given by

$$N \cdot r^{-r+o(r)}, \qquad \text{where} \quad \lim_{N\to\infty} \frac{o(r)}{r} = 0.$$

Let

$$\psi(x, y) = |\{m \in \mathbb{N}_{>0} \mid m \leq x \text{ and } m \text{ is } y\text{-smooth }\}|.$$

Then the theorem says that if $u = \frac{\ln x}{\ln y}$, that $\psi(x, y) \approx x u^{-u(1+o(1))}$ uniformly for $x \to \infty$ if there is a fixed $\varepsilon \in]0, 1[$ such that $(\ln x)^\varepsilon < u < (\ln x)^{1-\varepsilon}$.

For example if $x = 10^{100}$, $y = 10^{10}$, then $u = 10$ and $\psi(x, y) \approx 10^{100} \cdot 10^{-10} = 10^{90}$. This means that the probability that a random selected 100 digits number is 10^{10}-smooth is about 10^{-10}.

An improvement is to instead looking at random x_i to look at $x_i = a + \lfloor\sqrt{n}\rfloor\, b$ for small a, b, since

$$x_i^2 = a^2 + 2ab\,\lfloor\sqrt{n}\rfloor + b^2\,\lfloor\sqrt{n}\rfloor^2 \approx a^2 + 2ab\sqrt{n} + b^2n^2 \equiv a^2 + 2ab\sqrt{n} \pmod{n}.$$

Define the polynomial $q = (x + \lfloor\sqrt{an}\rfloor)^2 - an \in \mathbb{Z}[x]$, and assume that for some i and natural numbers k, $\bar{x}$ one has that p_i^k divides $q(\bar{x})$. Then p_i^k also divides $q(\bar{x} + \lambda p_i^k)$ for all $\lambda \in \mathbb{Z}$.

This allows to set up an efficient "sieve." For this fix a small $a \in \mathbb{N}$. The goal is to find all $x \in [0, I]$ for some $I > 0$ such that $q(x)$ is p_m-smooth. Set up an array

x	0	1	2	3	4	$\cdots$	$\bar{x}$	$\cdots$	$\bar{x} + p_i$	$\cdots$	$\cdots$	I
value adder	0	0	0	0	$\cdots$	0	$\log p_i^k$	0	$\log p_i^k$	0	$\cdots$	0

For different primes $p_i \in \{p_1, \ldots, p_m\}$ and natural numbers k solve $q(x) \equiv 0 \pmod{p_i^k}$. Let $\bar{x}$ be a solution. Add to "value adder" at the locations of the numbers $\hat{x} + \lambda p_i^k$, $\lambda \geq 0$, the value $\log p_i^k$. If for some number $v \in [0, I]$ the values add up to about $\log(2v\sqrt{an})$, then $q(v) = (v + \lfloor\sqrt{an}\rfloor)^2 - an$ is p_m-smooth.

A remark about the complexity of the quadratic sieve: one says a number theoretic problem has *subexponential running time* if there are numbers $\alpha \in]0, 1[$ and $c > 0$ such that the number of bit operations is

$$L_n(\alpha, c) := \mathcal{O}\Big(e^{c(\log n)^\alpha (\log\log n)^{1-\alpha}}\Big).$$

With $\alpha = 0$ this reduces to $\mathcal{O}((\log n)^c)$, i. e. polynomial time, and with $\alpha = 1$ to $\mathcal{O}(n^c$, i. e. exponential time. A careful analysis shows that the quadratic sieve's complexity is of the form $L_n(\frac{1}{2}, 1)$. The (currently) best known algorithm is an improvement of the quadratic sieve called the *generalized number field sieve*, which has $L_n(\frac{1}{3}, c)$ for $c \approx 1.92$ (???).

1.13.2 The Factorization Method of Claus Schnorr (1993)

The *basic idea* is to search for numbers of the form $p_1^{e_1} \cdots p_m^{e_m} \mod n$ which are "small". In other words: the resulting number is likely to be p_m-smooth. If this is possible, one would have relations of the form

$$\prod_{i=1}^{m} p_i^{e_i} \mod n = \prod_{i=1}^{m} p_i^{\tilde{e}_i},$$

and if one has more than m such relations an equation $x^2 \equiv y^2 \pmod{n}$ can be constructed. Write $N = \prod p_i^{e_i} \approx \prod p_i^{f_i} = \hat{N}$. We want that $\left|N - \hat{N}\right| < s$, where s is a number with the property that with high probability $\left|N - \hat{N}\right|$ is p_m-smooth. (Note that if $N - \hat{N}$ is p_m-smooth, then so are $N - \hat{N} \mod n$ and $\hat{N} - N \mod n$.)

Taking logarithms we get

$$(e_1 - f_1)\log p_1 + \cdots + (e_m - f_m)\log p_m \approx \log n.$$

The requirement $\left|N - \hat{N}\right| < s$ translates into

$$\left|\sum_{i=1}^{m} (e_i - f_i)\log p_i - \log n\right| \leq \frac{1}{N} s$$

(which can be get by Taylor expansion). Consider the *Schnorr lattice*, which is the lattice Λ defined by the matrix

$$M = \left(\begin{array}{ccc|c} \log p_1 & & 0 & N\log p_1 \\ & \ddots & & \vdots \\ 0 & & \log p_m & N\log p_m \\ \hline 0 & \cdots & 0 & N\log n \end{array}\right) \in \mathbb{R}^{(m+1)\times(m+1)}.$$

Consider a linear combination

$$(\lambda_1, \ldots, \lambda_m, c)M = (\lambda_1 \log p_1, \ldots, \lambda_m \log p_m, \sum \lambda_i N \log p_i + cN \log n).$$

(It is possible to round all entries to the nearest integer and get an integer lattice.)

Theorem 1.13.2 (C. Schnorr). *Let $c > 1$ be a fixed real number and $N = n^c$, where n is the value to be factored. If $(\lambda_1, \ldots, \lambda_m) \in \mathbb{Z}^m$ satisfy the inequalities*

(i) $|\sum_{i=1}^m \lambda_i \log p_i - \log n| \leq \frac{1}{N} p_m$ *and*

(ii) $\sum_{i=1}^m |\lambda_i \log p_i| < (2c-1)\log n + 2\log p_m$,

then for

$$u = \prod_{i=1}^m p_i^{e_i} \qquad \text{and} \qquad v = \prod_{i=1}^m p_i^{f_i}$$

we have $|u - vn| \leq p_m^2$.

In his 1993 paper Schnorr estimated that for factoring a number with 512 bits, a lattice of size $m = 6300$ should be reduced with an algorithm more costly than LLL (namely the Korkin-Zolotoa algorithm).

1.13.3 Lenstras Elliptic Curve Factorization Method

We first recall Pollards $(p-1)$-method. Assume $n = pq$ and for some bound $B > 0$ it is that $p - 1$ is B-smooth, while $q - 1$ is not. Let

$$k = \prod_{\substack{u \leq B \\ u \text{ prime}}} u^{\left\lfloor \frac{\log n}{\log u} \right\rfloor}.$$

(Then $p - 1$ divides k, but $q - 1$ does not.) For all $x \in \mathbb{Z}_p^*$ we have $x^k = 1$ in $\mathbb{Z}_p$. Also $\{x \in \mathbb{Z}_q^* \mid x^k = 1 \in \mathbb{Z}_q\}$ is a proper subgroup of $\mathbb{Z}_q^*$. Taking a random number $x \in \mathbb{Z}_n^*$, with probability at least 50 % we have that

$$\gcd(x^k - 1 \mod n, n) \in \{p, n\}.$$

(The gcd cannot be 1 or q.)

Remarks 1.13.3.

(1) *Computation of $x^k - 1 \mod n$ can be done iteratively: start with $x_0 := x$ and compute*

$$x_{i+1} := x_i^{p_i^{\left\lfloor \frac{\log n}{\log p_i} \right\rfloor}} \mod n.$$

Then test if $\gcd(x_{i+1} - 1 \mod n, n) \neq 1$. (Here $p_1, p_2, \ldots$ denote the first primes and x is a random number in $\mathbb{Z}_n$.)

(2) *If both $p - 1$ and $q - 1$ have large prime factors, this method will fail.*

(3) *Abstractly the $(p-1)$-algorithm is based on group homomorphisms $\varphi : \mathbb{Z}_n^* \to \mathbb{F}_p^*$ (reduction modulo p), and one searches for a kernel element.*

Lenstras Method Consider an elliptic curve $E : y^2 = x^3 + ax + b$, where $a, b \in \mathbb{Z}$. Assume $P = (\alpha, \beta) \in \mathbb{Z}^2$ is a point on the curve, and assume $n = \prod_{i=1}^m p_i^{e_i}$ where the p_i are distinct primes and $e_i \in \mathbb{N}_{>0}$. We can look at the curve in different ways:

- As a curve over $\mathbb{Q}$;
- As a curve over $\mathbb{Z}_{p_i}$;
- As a curve over $\mathbb{Z}_n$, which by reduction modulo p_i reduces to a curve over $\mathbb{Z}_{p_i}$.

To exploit the group laws for the first two kinds of curves, we require that $\Delta = 4a^3 + 27b^2 \in \mathbb{Z}_n^*$. It is easy to construct such a curve with a point on it: pick randomly integers $a, \alpha \in \mathbb{Z}_n$, and then choose b such that $\alpha^3 + a\alpha + b$ is a square in $\mathbb{Z}$, like β^2. Then $y^2 + x^3 + ax + b$ contains $P = (\alpha, \beta) \in \mathbb{Z}_n^2$. For smoothness compute $\gcd(\Delta \mod n, n)$; it is either 0, 1 or any other positive number. If it is 1 we are done, if it is 0 we restart, and otherwise we found a non-trivial factor of n.

Let $\ell_1, \ldots, \ell_m$ be the orders of P modulo p_i, i. e. in the groups $E(\mathbb{F}_{p_i})$. Assume that ℓ_i is B-smooth and another ℓ_j is not. As in Pollards $(p-1)$-method, compute

$$k := \prod_{\substack{u \le B \\ u \text{ prime}}} u^{\left\lfloor \frac{\log n}{\log u} \right\rfloor}$$

and compute $kP \in E(\mathbb{Z}_n)$ using the normal addition formulas. If ℓ_i is B-smooth, then $kP = \mathcal{O}$ in $E(\mathbb{F}_{p_i})$, and if ℓ_j is not B-smooth, then $kP \neq \mathcal{O}$ in $E(\mathbb{F}_{p_j})$. If we work in homogenous coordinates $kP = (x : y : z)$, we have $z \equiv 0 \pmod{p_i}$, but $z \not\equiv 0 \pmod{p_j}$—thus we have found a non-trivial factor of n!

Remarks 1.13.4.

(1) *In order to compute $Q = kP \in E(\mathbb{Z}_n)$ we treat $\mathbb{Z}_n$ as if it would be a field and use the usual addition formulas. If at some step inversion modulo n is not possible though it should be, we have a factor. Thus we do not have to work in homogenous coordinates, but can also use inhomogenous coordinates.*

(2) *As in Pollard $(p-1)$ we can compute kP iteratively.*

The Complexity Let $n = \prod_{i=1}^m p_i^{e_i}$ be as above, where $p_1 < \cdots < p_m$. The algorithm succeeds as soon as some of the orders $\ell_1, \ldots, \ell_m$ of P in $E(\mathbb{F}_{p_i})$ are B-smooth while others are not. The expected size of ℓ_i is $|E(\mathbb{F}_{p_i})|$, which is the size of p_i by Hasse. Using some heuristic arguments Lenstra computed the asymptotic complexity as

$$L_p(\tfrac{1}{2}, \sqrt{2}) = \mathcal{O}\Big(\exp \sqrt{(2 + o(1))(\log p)(\log\log p)} \cdot (\log n)^2\Big).$$

For the quadratic sieve one gets

$$L_n(\tfrac{1}{2}, 1) = \mathcal{O}\Big(\exp \sqrt{(1 + o(1))(\log n)(\log\log n)}\Big).$$

In situations where $n = pq$ and $p \approx q$, $p \neq q$, we see that asymptotically $L_n(\frac{1}{2}, 1) \approx L_p(\frac{1}{2}, \sqrt{2})$, thus the algorithms have asymptotically the same complexity.

Caveat The arithmetic operations per addition in Lenstras algorithm are much more costly!

Remark 1.13.5. *The quadratic sieve can be generalized from $\mathbb{Q}$ to number fields, i. e. finite extensions of $\mathbb{Q}$, where one works for example in $\mathbb{Q}[\sqrt{3}]$ which is as a $\mathbb{Q}$-vector space isomorphic to $\mathbb{Q} \oplus \sqrt{3}\mathbb{Q}$. An idea would be to do a similar generalization with Lenstra's algorithm.*

1.14 Hash Functions

Let X and Y be sets, where Y is finite and X is possibly infinite.

Definition 1.14.1. *A one-way function $h : X \to Y$ is called a* (cryptographic) hash function.

This means that for a given $y \in Y$ it is computationally not feasible to find an $x \in X$ such that $h(x) = y$.

Remark 1.14.2. *In practice X is typically of the form*

$$X = A^* := \bigcup_{i=0}^{\infty} A^i,$$

where A is some (finite) alphabet.

Applications 1.14.3.

(1) *Simple error protection.*

(2) *In connection with digital signatures:*

If an email should be signed, one in practice computes the hash of the mail and includes that hash value in the signature, so that anyone can check whether the signature (if it is valid) belongs to the mail.

Definition 1.14.4. *A hash function is called* weakly collision free *if for a particular $x \in X$ it is computationally not feasible to find an $x' \in X$ such that $h(x) = h(x')$ and $x' \neq x$. It is called* strongly collision free *if it is not feasible to find two distinct elements $x, x' \in X$ such that $h(x) = h(x')$.*

Remark 1.14.5. *If h is weakly collision free, and the values are uniformly distributed, then finding an x' through a random search requires $\mathcal{O}(|Y|)$ trials. For strongly collision free hash functions the number of trials is $\mathcal{O}(\sqrt{|Y|})$.*

As an example, consider the MD5 hash function. (Note that recently collisions were found for MD5.)

1.14.1 The Chaum-van Heijst-Pfitzmann Hash Function

Let p, q be distinct primes with $p = 2q + 1$ (i. e. p is a safe prime; recall that such primes are easy to find) Let α and β be two primitives of $\mathbb{F}_p = \mathbb{Z}_p$. (Actually we have that $(\mathbb{Z}_p, \cdot) \cong (\mathbb{Z}_{2q}, +) \cong (\mathbb{Z}_q, +) \oplus (\mathbb{Z}_q, +)$, and thus half of the units of $\mathbb{Z}_p$ are primitive.) Identify $\mathbb{F}_q = \mathbb{Z}_q$ with $\{0, 1, \ldots, q-1\} \subseteq \mathbb{N}$.

Lemma 1.14.6. *The function $h : \mathbb{F}_q \times \mathbb{F}_q \to \mathbb{F}_p^*$, $(x_1, x_2) \mapsto \alpha^{x_1}\beta^{x_2}$ can serve as a hash function: finding a collision is equivalent with solving $\log_\alpha \beta$.*

Proof. Assume $s = \log_\alpha \beta$ is known. Then $\alpha^{x_1}\beta^{x_2} = \alpha^{x_1+s}\beta^{x_2-1}$, and therefore (x_1, x_2) and $(x_1 + s, x_2 - 1)$ hash to the same value.

Vice versa: let $(x_1, x_2) \neq (x_3, x_4)$ be pairs such that $h(x_1, x_2) = h(x_3, x_4)$. Thus $\alpha^{x_1-x_3} = \beta^{x_4-x_2}$. If $x_4 = x_2$, then $x_1 = x_3$; therefore we can assume $x_1 \neq x_3$, $x_2 \neq x_4$. Without loss of generality $x_4 > x_2$. Let $d = \gcd(x_4 - x_2, p - 1)$. Since $q > x_4 - x_2 \geq 1$ and $p - 1 = 2q$, either $d = 1$ or $d = 2$.

If $d = 1$, then with $y := (x_4 - x_2)^{-1} \pmod{p-1}$ we have $(\beta^{x_4-x_2})^y = \beta$, and moreover $(\beta^{x_4-x_2})^y = (\alpha^{x_3-x_1})^y = \alpha^{(x_3-x_1)y}$ and thus $(x_3 - x_1)y = \log_\alpha \beta$.

If $d = 2$ we have $\gcd(x_4 - x_2, q) = 1$ since $p - 1 = 2q$. Let $y := (x_4 - x_2)^{-1} \pmod{q}$. Write $y(x_4 - x_2) = 1 + kq$, $k \in \mathbb{Z}$. Then we have

$$\beta^{qk}\beta = \beta^{qk+1} \cong \beta^{y(x_3-x_1)} \pmod{p},$$

and $\beta^{kq} \in \{-1, 1\}$. (In fact we have $\beta^{kq} \cong (-1)^k$.) Therefore we also get $\log_\alpha \beta \in \{(x_4 - x_2)y, (x_4 - x_2)y + q\}$, and a simple trial reveals $\log_\alpha \beta$. □

1.14.2 Construction of Practical Hash Functions

Question 1.14.7. *Given a one-way function $h : X \to Y$, both X and Y finite. How to construct a hash function*

$$h^* : X^* \to Y, \qquad \text{where } X^* = \bigcup_{i=0}^{\infty} X^i.$$

Method #1 Assume $X = Y$ and that X has some additive structure. For each n, define a function $h_n : X^n \to Y$, $(x_1, \ldots, x_n) \mapsto h_n(x_1, \ldots, x_n)$ and $h^* : X^* \to Y$, $(x_1, \ldots, x_n) \mapsto h_n(x_1, \ldots, x_n)$. The function h_n is defined recursively by

$$h_{n+1}(x_1, \ldots, x_n, x_{n+1}) := h(x_{n+1} + h_n(x_1, \ldots, x_n) \qquad \text{and} \qquad h_1(x_1) := h(x_1).$$

Method #2 Based on a secret key system $f : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$, where $\mathcal{M} \cong \mathcal{K} \cong \mathcal{C}$ as sets, define $X := \mathcal{K}$ and $Y := \mathcal{C}$. From the definition of a secret key system we know that for any $m \in M$ the function $x \mapsto f(m, x)$ is a one-way and hence a hash function. Given $(x_1, \ldots, x_n) \in X^n$, define $h^n : X^n \to Y$ through the recurrence relation $y_1 := m$, $y_{i+1} := f(y_i, x_i)$ and $h^n(x_1, \ldots, x_n) := y_{n+1}$. As in method #1, this defines a function $h^* : X^* \to Y$.

As an *exercise*, assume that $f : \mathcal{M} \times \mathcal{K} \to \mathcal{C}$ is a secret key system such that $f_m : \mathcal{K} \to \mathcal{C}$ is strongly collision free for every $m \in \mathcal{M}$. Show that $h^* : X^* \to Y$ is strongly collision free, or prove that it is not.

1.15 Protocols

1.15.1 Secret Sharing Systems

Consider the following situation: a bank wants to give to N employees access to a tresor in a way where n of them can open the tresor together, but less than n can not.

Repetition of Lagrange Interpolation Let $\mathbb{F}$ be a field and $\{(x_0, y_0), \dots, (x_n, y_n)\} \subseteq \mathbb{F}^2$ be $(n+1)$ points with $x_i \neq x_j$ for $i \neq j$.

Lemma 1.15.1. *There exists a unique polynomial $f \in \mathbb{F}[x]$ of degree n such that $f(x_i) = y_i$ for every i.*

Proof. Assume $\tilde{f} \in \mathbb{F}[x]$ is another such polynomial. Then $f - \tilde{f}$ is a polynomial with $n+1$ roots (in $x = x_i$), and therefore the zero polynomial, and thus $f = \tilde{f}$.

For the existence define $f_i := \prod_{j \neq i} \frac{x - x_j}{x_i - x_j} \in \mathbb{F}[x]$. We have $\deg f_i = n$ and $f_i(x_i) = 1$, $f_i(x_j) = 0$ for $j \neq i$. Therefore $f = \sum y_i f_i$ is the required polynomial. □

Shamir Treshold Scheme Assume N employees should be able to access the tresor if and only if at least$n$ of them are present. Choose a finite field $\mathbb{F}$ with $|\mathbb{F}| \geq 2^{60}$, and choose a random polynomial $f = \sum_{i=0}^{n-1} a_i x^i \in \mathbb{F}[x]$, $a_{n-1} \neq 0$, and choose random $x_1, \dots, x_N \in \mathbb{F}^*$ where $x_i \neq x_j$ for $i \neq j$. Each employee receives a personal partial key $(x_i, f(x_i)) \in \mathbb{F}^2$. The secret key of the tresor is $f(0) = a$.

Any n employees $1 \leq i_1 < \cdots < i_n \leq N$ can compute f by Lagrange interpolation using the $(x_{i_j}, f(x_{i_j}))$ pairs, $j = 1, \dots, n$, and therefore $f(0) = a$.

Remark 1.15.2. *Any $n-1$ employees or less have zero knowledge of $f(0)$, since for every $\hat{a} \in \mathbb{F}$ there exists an $\hat{f} \in \mathbb{F}[x]$ of degree $n-1$ such that $\hat{f}(0) = \hat{a}$ and $\hat{f}(x_i) = f(x_i)$ for less than n of the i's.*

1.15.2 Signature Schemes

Let $\mathcal{M}$ be a set of message words, and $\mathcal{S}$ a set of possible signatures. A *signature scheme* consists of a (secret) signing function $sign : \mathcal{M} \to \mathcal{S}$ and a publicly known verification function $verify : \mathcal{M} \times \mathcal{S} \to \{true, false\}$ such that $verify(m, s) = true$ if and only if $sign(m) = s$ for $s \in \mathcal{S}$, $m \in \mathcal{M}$.

Remark 1.15.3. *For the forger Oscar it should not be possible to construct a pair $(m, s) \in \mathcal{M} \times \mathcal{S}$ such that $verify(m, s) = true$.*

Example 1.15.4. *Based on RSA: the public function is $\psi : \mathbb{Z}_n \to \mathbb{Z}_n$, $x \mapsto x^e$, and the private function is $sign : \mathbb{Z}_n \to \mathbb{Z}_n$, $x \mapsto x^d$. Then $verify(m, s) = true$ if and only if $\psi(s) = m$. Here $\mathcal{M} = \mathcal{S} = \mathbb{Z}_n$.*

Note that for this scheme Oscar can start with an $s \in \mathcal{S}$ and generate m such that (m, s) is correct!

A solution to this problem is the following scheme: the scheme should be set up such that a random $m \in \mathbb{Z}_n$ is not a valid message to be signed. One way to accomplish this is via a hash function: has the message to be signed, sign the hash value and send both signature and message.

ElGamal Signature Scheme The ElGamal signature scheme dates back to 1985. ElGamal proposed the following scheme:

Let p be a prime, $p \approx 2^{1000}$, and $\alpha \in \mathbb{Z}_p^*$ a primitive element and $\beta = \alpha^a$ for some $a \in \mathbb{Z}_{p-1}$ which is only known to the signer. The sign function is

$$sign : \mathbb{Z}_p \to \mathbb{Z}_p \times \mathbb{Z}_{p-1}, \qquad m \mapsto (\alpha^k, (m - a\alpha^k)k^{-1}) = (s_1, s_2);$$

here k is randomly chosen. In this formula, α^k should be computed in $\mathbb{Z}_p$ in both components, while the other operations in the second component should be computed in $\mathbb{Z}_{p-1}$. The signature consists of (m, s_1, s_2).

Public data are α, β and p, which are deposited with a trusted authority. Private data is $a = \log_\alpha \beta$. To verify, note that we have

$$\beta^{s_1} s_1^{s_2} \equiv \alpha^{a\alpha^k} \alpha^{k(m-a\alpha^k)k^{-1}} \equiv \alpha^{a\alpha^k + m - a\alpha^k} \equiv \alpha^m \pmod{p}.$$

Assume that Oscar wants to forge a signature in order to sign a message m. He has to find $(s_1, s_2) \in \mathbb{Z}_p \times \mathbb{Z}_{p-1}$ such that $\beta^{s_1} s_1^{s_2} \equiv \alpha^m \pmod{p}$. If he randomly choses s_1, he has to solve the DLP $s_1^{s_2} \equiv \alpha^m \beta^{-s_1}$, i. e. $s_2 = \log_{s_1}(\alpha^m \beta^{-s_1})$.

The drawback of this scheme is that if $p \approx 2^{1000}$, the signature requires 3000 bits of data storage.

The Digital Signature Algorithm (DSA) In 1994 the National Institute for Standards in Technology (NIST) adopted a variation of the ElGamal signature scheme as the standard called Digital Signature Algorithm (DSA).

For this scheme $p \approx 2^{1000}$, and we have the functions

$$\begin{aligned} sign &: \mathbb{Z}_p \to \mathbb{Z}_p \times \mathbb{Z}_{p-1}, \\ &\quad m \mapsto (\alpha^k, (m + a\alpha^k)k^{-1}) = (s_1, s_2); \\ verify &: \mathbb{Z}_p \times \mathbb{Z}_p \times \mathbb{Z}_{p-1} \to \{true, false\}, \\ &(m, (s_1, s_2)) \mapsto \begin{cases} true & \text{if } \alpha^m \beta^{s_1} \equiv s_1^{s_2} \pmod{p}, \\ false & \text{otherwise.} \end{cases} \end{aligned}$$

If s_2 is invertible the equation is equivalent to

$$\alpha^{m s_2^{-1}} \beta^{s_1 s_2^{-1}} \equiv s_1 \pmod{p}.$$

Let q be a second prime such that q divides $p-1$, and let $\alpha_0 = \alpha^{\frac{p-1}{q}} \mod q$ and $\beta_0 = \alpha_0^a \mod q = \beta^{\frac{p-1}{q}} \mod q$. Then both α_0, β_0 are q-th roots of unity. We can ask whether

$$\alpha_0^{m s_2^{-1}} \beta_0^{s_1 s_2^{-1}} \equiv s_1^{\frac{p-1}{q}} \pmod{p}.$$

Let $\tilde{m} = m \mod q$, $\tilde{s}_1 = s_1 \mod q$ and $\tilde{s}_2 = s_2 \mod q$. Then $(\tilde{m}, \tilde{s}_1, \tilde{s}_2)$ will serve as the signature, and verification goes by

$$\alpha_0^{\tilde{m}\tilde{s}_2^{-1}} \beta_0^{\tilde{s}_1 \tilde{s}_2^{-1}} \equiv \left(\tilde{s}^{\frac{p-1}{q}} \mod p\right) \pmod{q}.$$

Note that $\tilde{m}$, $\tilde{s}_1$ and $\tilde{s}_2$ all have a simliar size than q. The adopted DSA standard specifies:

- The person who wants to have a signature function chooses $2^{159} \le q \le 2^{160}$, q prime, and searches for a prime p, $2^{512} \le p \le 2^{524}$ such that q divides $p-1$. (This is easily accomplished, just search for primes of the form $\gamma q + 1$.)
- Select a primitive $\alpha \in \mathbb{Z}_p^*$ and let$\alpha_0 := \alpha^{\frac{p-1}{q}} \pmod{p}$.
- Select $0 < a < q$ randomly and let $\beta_0 := \alpha_0^a \pmod{p}$.
- The public data ($\to$ trusted authority) are α_0, β_0, p and q.
- The sign function is defined as

$$\begin{aligned} sign &: \mathbb{Z}_q \to \mathbb{Z}_q \times \mathbb{Z}_q, \\ &\quad m \mapsto ((\alpha_0^k \mod p) \mod q, ((m + a\alpha_0^k)k^{-1} \mod p) \mod q), \end{aligned}$$

  and the verification function as

$$\begin{aligned} verify &: \mathbb{Z}_q^3 \to \{true, false\}, \\ &(m, (s_1, s_2)) \mapsto \begin{cases} true & \text{if } \alpha_0^{m s_2^{-1}} \beta_0^{s_1 s_2^{-1}} \mod p \equiv s_1^{\frac{p-1}{q}} \mod p \pmod{q}, \\ false & \text{otherwise.} \end{cases} \end{aligned}$$

  The total signature is about $3 \cdot 160 = 480$ bits.

1.15.3 Identification Schemes

The goal is that one party ('the verifyer') can make sure that 'the claimant' is the person he/she claims to be. This occurs in several practical situations:

(a) Over the internet, how does a bank knows that it is Alice that tries to access her account?

(b) How does Alice knows it is the bank she is dealing with? (Recently there have been lot of attacks known as *phishing* where people should be lured to websites looking like their bank's website.)

Question (a) is usually dealt with using a (one-time) password. Question (b) is usually dealt with using an identification protocol involving a trusted party; examples for 'trusted parties' are companies like verisign and entrust.

The Fiat-Shamir Protocol Recall: if $n = pq$ where p and q are distinct primes, we have that the following problems are equivalent:

(a) Finding all four solutions of a random quadratic equation $x^2 + bx + c \equiv 0 \pmod{n}$;

(b) Computing all four solutions of $z^2 + (c - b^2/4) = 0$ in $\mathbb{Z}_n$;

(c) Factoring n.

The Fiat-Shamir scheme works as follows:

A trusted authority (TA) chooses $n = pq$ and keeps p and q secret (or even destroys them). The bank (B) registers with the TA by choosing a random integer $s \in \mathbb{Z}_n$ and by computing $v = s^2 \in \mathbb{Z}_n$. The TA keeps a file

$$\text{Bank} \leftrightarrow v.$$

Assume that the bank wants to identify itself to Alice. For this the bank will convince Alice that they know s using a *zero-knowledge proof*: the bank chooses a random number $r \in \mathbb{Z}_n$ and transmits to Alice $x := r^2 \in \mathbb{Z}_n$. Alice challenges the bank by asking one of the two questions:

(a) Compute $y = rs$. This can be verified by Alice through $y^2 = r^2 s^2 = vx$.

(b) What is r? Alice verifies the answer by computing $r^2 = x$.

Of course Alice can not ask both questions, as this would reveal s. Note that if Olga tries to impersonate the bank, she can choose $r \in \mathbb{Z}_n$, compute $x = r^2$ what allows her to answer (b), but not (a). Alternatively she can choose r and compute $\tilde{x} = r^2/v = r^2/s^2 = \tilde{r}^2$. If Alice asks for $\tilde{r}s$, Olga provides an answer for (a), but not for (b). (Indeed $\tilde{x}^2 s^2 = r^2$.)

The Schnorr Identification Scheme The TA generates a prime $q \geq 2^{140}$ and a prime p such that q divides $p - 1$. (The same setup as for DSA.) Let $\alpha \in \mathbb{Z}_p^*$ be a primitive element and $\alpha_0 = \alpha^{\frac{p-1}{q}} \mod p$ a q-th root of unity. The *public data* are α_0, p and q. Assume Alice wants to register with the TA. She chooses a random integer e, $0 < e < q$, keeps e secret and gives the TA the number $v = \alpha_0^{-e} \mod p$. The TA publishes for every user a number

$$v \leftrightarrow \text{Alice}.$$

Now assume Alice wants to identify herself to Bob. She chooses a random integer k, $1 \leq k < q$, and sends $\gamma = \alpha_0^k \mod p$ to Bob. Bob chooses a random integer r, $1 \leq r < q$, and gives it to Alice. Alice computes $y = k + er \mod q$ and gives it back to Bob. Now Bob verifies that

$$\alpha_0^y v^r \equiv \alpha_0^{k+er} \alpha_0^{-er} \equiv \alpha_0^k \equiv \gamma \pmod{p}.$$

The *security* lies in the fact that only Alice knows k and e, and somebody impersonating her has to compute e from $y = k + er \in \mathbb{Z}_q$ (one equation in two unknowns), $\alpha_0^{-e} = v \in \mathbb{Z}_p$ (a DLP in $\mathbb{Z}_p$).

The underlying principle is once more a *zero-knowledge proof*: Alice proves to Bob that she knows $-\log_\alpha v = e$ without revealing anything about e.

Bibliography

[AKS02] Manindra Agrawal, Neeraj Kayal, and Nitin Saxena. PRIMES is in P. Avaible online on `http://www.cse.iitk.ac.in/news/primality.html`, August 2002.

[MvOV96] Alfred Menezes, P. van Oorschot, and S. A. Vanstone. *Handbook of Applied Cryptography*. CRC Press Series on Discrete Mathematics and its Applications. CRC Press, Boca Raton, FL, 1996.

Index