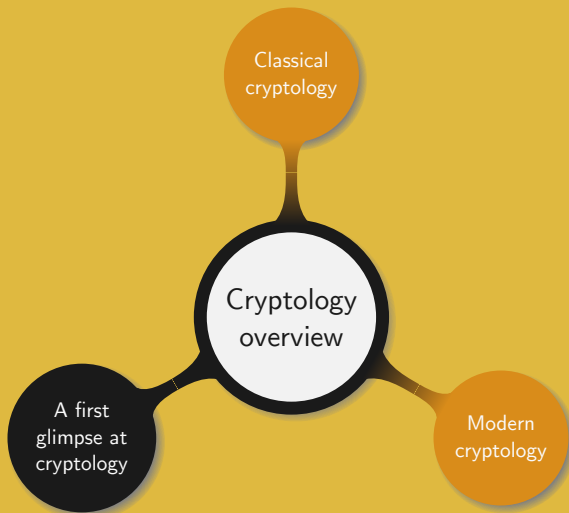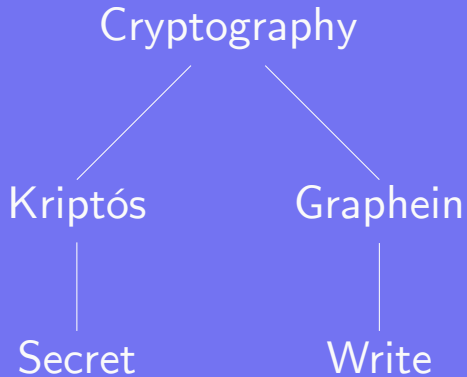# Introduction to Cryptography
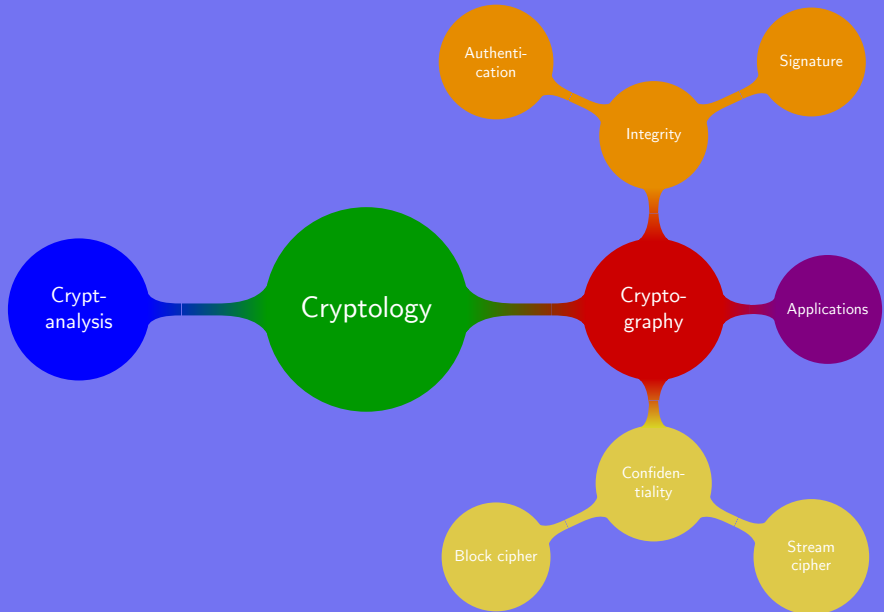
1. Cryptology overview

Manuel – Summer 2022

Are you following the right course?

Cryptography

Kriptós   Graphein

Secret   Write

Cryptology as an old science:

-1500	Mesopotamia: earliest known use
-700	Greeks: military applications
-500	Common among Hebrew scholars
-100	Romans
800	Arabs: document cryptanalysis methods
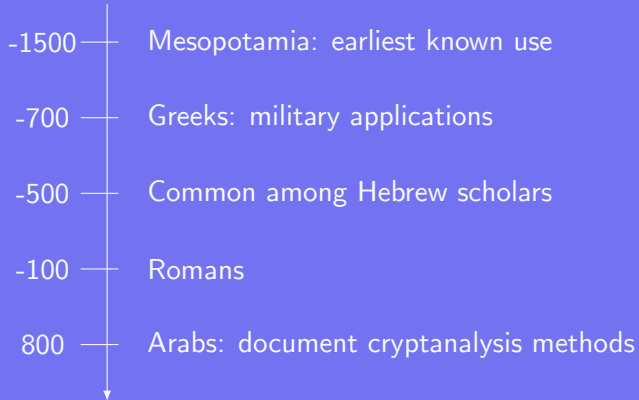
Advantage was on the side of cryptanalysts

No major advances until World War I:

1914	Basic and insecure encryption methods
1917	Discovery of an unbreakable cipher
1920	Development of electromechanical devices
1929	First usage of mathematics
1939	Major breakthroughs
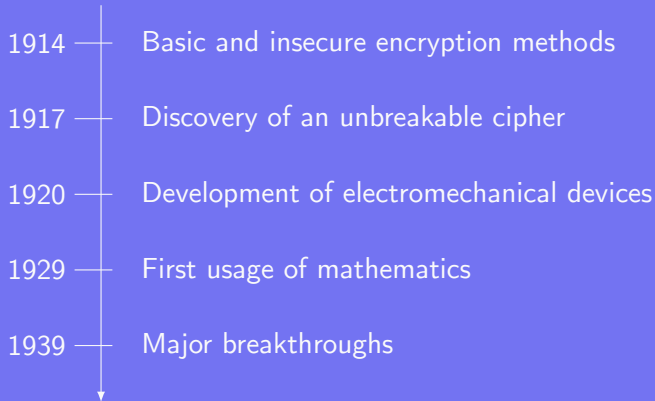
Advantage is still on the side of cryptanalysts

Mathematics becomes the heart of cryptography:

1949	Shannon: secrecy et authenticity
1976	Diffie et Hellman: public key cryptography
1990	Internet popularize the use of cryptography
2017	Cryptology is everywhere
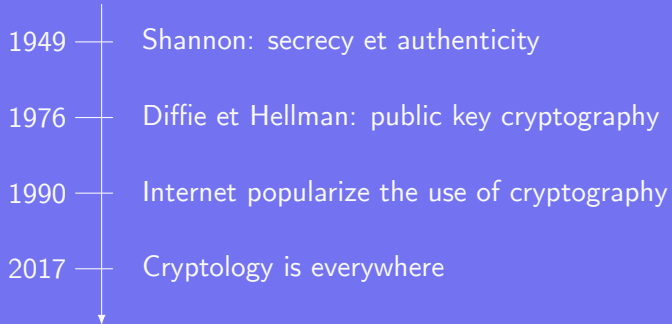
Politics try to kill cryptography and give governments the monopoly
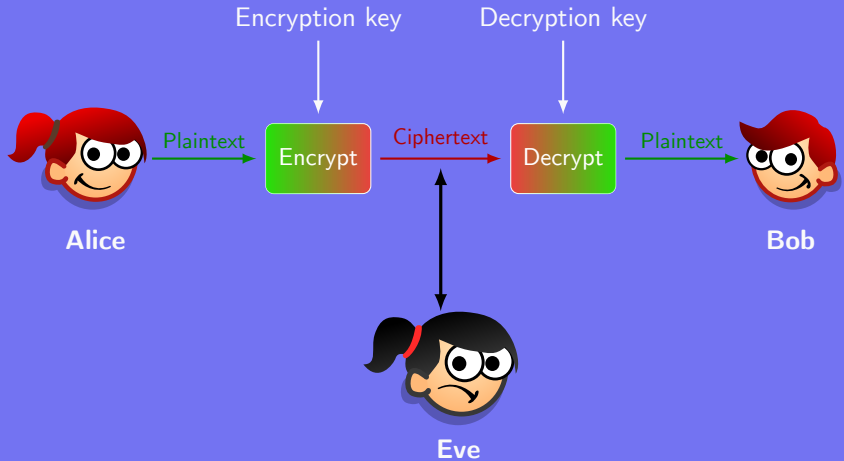
Eve has one of the following goals:

- Read a message

- Find the key

- Corrupt Alice's message

- Masquerade as Alice

There are the five main types of attacks:

- Eve only has a copy of the ciphertext: *ciphertext only*

- Eve has a copy of the ciphertext but also of the corresponding plaintext: *Known Plaintext Attack (KPA)*

- Eve chooses the plaintext to be encrypted: *Chosen Plaintext Attack (CPA)*

- Eve chooses the ciphertext to be decrypted: *Chosen Ciphertext Attack (CCA)*

- Eve chooses any plaintext to be encrypted or ciphertext to be decrypted: *Chosen Plaintext and Ciphertext Attack (CPCA)*

Methods to collect data:

- On fiber cables and infrastructures as the flow passes

- From the servers of service providers

Methods to retrieve encrypted data:

- Break the encryption

- Influence industrial standards

- Pressure manufacturers to make insecure devices

- Infiltrate hardware and software

Eve is anyone that might want to read or temper the data:

- Low threat: friends, family members, etc.

- High threat: governmental agencies and companies

Reasons for mass surveillance:

- Combat terrorism

- Assess foreign policies and economical stability

- Gather commercial secrets

What does your phone know about you?

"They (the NSA) can use the system to go back in time and scrutinize every decision you've ever made, every friend you've ever discussed something with, and attack you on that basis to sort of derive suspicion from an innocent life and paint anyone in the context of a wrongdoer."
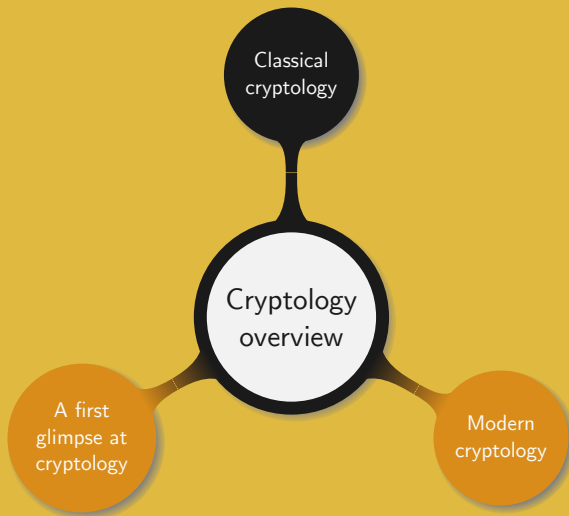
Edward Snowden

Principle (Kerckhoffs' principle)

A cryptosystem should be secure even if everything about the system, except the key, is public knowledge.

In other words:

- Security through obscurity is not security

- Data should be encrypted using standard, publicly known algorithms

- The implementation must be accessible to all
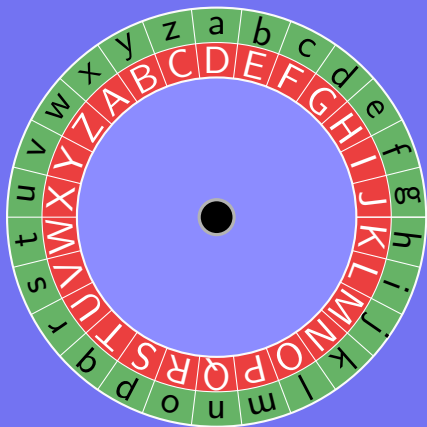
Simple description:

- One of the earliest cipher

- Attributed to Julius Caesar

- Letters are shifted by a given number of places

- The shift is called the *key* of the cipher

Exercise. Recover the plaintext given the ciphertext JRRGPRUQLQJ

Definitions

1. Let a and b be two integers, with $a \neq 0$. We say that a *divides* b if there exists an integer k such that $b = ak$, and we denote it $a|b$.

2. Let a, b and n be three integers with $n \neq 0$. We say that a *is congruent to* b *modulo* n, if n divides $a - b$. It is denoted $a \equiv b \bmod n$

In modern cryptography:

- The plaintext is first converted into a numerical value

- If the alphabet is composed of n symbols then each one is assigned a value between 0 and $n - 1$

Caesar cipher in mathematical terms:

1. Label letters as integers from 0 to 25

2. Choose a key κ in the range 0 – 25

3. Encrypt using the function $x \mapsto x + \kappa \bmod 26$

4. Decrypt using the function $x \mapsto x - \kappa \bmod 26$

5. Label integers from 0 to 25 as letters

Exercise. Encrypt and decrypt "students are working hard" using Caesar cipher with the key $\kappa = -5$

Using the different types of attacks:

- Ciphertext only: only 26 possible keys $\Rightarrow$ exhaustive search

- KPA: deduce the key from the plaintext/ciphertext pair

- CPA: for the plaintext "a", the ciphertext gives κ

- CCA: for the ciphertext "A", the plaintext gives $-\kappa \bmod 26$

In 1776 Thomas Jefferson sent a message to Benjamin Franklin:

LWNSOZBNWVWBAYBNVBSQWVUOHWDIZWRBBNPBPOOUWRPAWX
AWPBWZWMYPOBNPBBNWJPAWWRZSLWZQJBNVIAXAWPBSALIBNX
WABPIRYRPOIWRPQOWAIENBVBVBNPBPUSREBNWVVWPAWOIHWOIQW
ABJPRZBNWFYAVYIBSHNPFFIRWVVBNPBBSVWXYAWBNWVWAIENB
VESDWARUWRBVPAWIRVBIBYBWZPUSREUWRZWAIDIREBHWIATYVB
FSLWAVHASUBNWXSRVWRBSHBOTESDWARWZBNPBLNWWDWAPRJ
HSAUSHESDWARUWRBQWXSUWVZWVBAYXBIDWSHBNWVWVWWRZVIB
IVBNVAIENBSHBNWFWSFOWBSPOBSPQSOIVNIBPRZBSIRVBIBYVBI
BYBWRWLESDWARUWRBOPJIREIBVHSYRZPBISRSRVYXNFAIRXIFOO
TPRZSAEPRIKIREIBVFSLWAVIRVYXNHSAUPVBSVWMJSVBOICWOJBS
WHHWXBBNWIAVPHWBJPRZNPFFIRWW

Your job is to decipher it
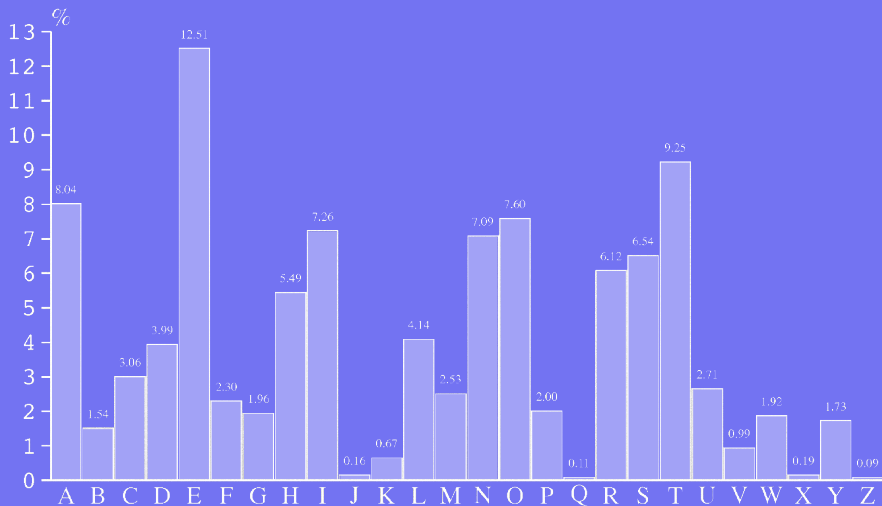
For the 10 most common letters their count gives:

W	B	R	S	I	V	A	P	N	O
76	64	39	36	36	35	34	32	30	16

We can guess:

- W is probably e
- B, R, S, I, V, A, P, and N are probably t, a, o, i, n, s, h, and r
- But what is their order?

Digrams count

	W	B	R	S	I	V	A	P	N
W	3	4	12	2	4	10	14	3	1
B	4	4	0	11	5	5	2	4	20
R	5	5	0	1	1	5	0	3	0
S	1	0	5	0	1	3	5	2	0
I	1	8	10	1	0	2	3	0	0
V	8	10	0	0	2	2	0	3	1
A	7	3	4	2	5	4	0	1	0
P	0	8	6	0	1	1	4	0	0
N	14	3	0	1	1	1	0	7	0

Rules in English

- e contacts most of other letters
- a, i, o tend to avoid each other
- 80% of the letters preceding n are vowels
- the most common digram is th
- h often appears before e, rarely after
- r pairs more with vowels and s with consonants
- rn more common than nr and to than ot

Summarizing all the guesses and carrying on:

L	W	N	S	O	Z	B	N	W	V	W	B	A	Y
w	**e**	**h**	**o**	**l**	**d**	t	h	e	s	e	t	r	u
B	N	V	B	S	Q	W	V	W	O	H	W	D	I
t	h	s	t	o	**b**	e	s	e	l	f	e	v	i
Z	W	R	B	B	N	P	B	P	...				
d	e	n	t	t	h	a	t	a					

The deciphered text is from the Declaration of independence:

we hold these truths to be self evident that all men are created equal that they are endowed by their creator with certain unalienable rights that among these are life liberty and the pursuit of happiness that to secure these rights governments are instituted among men deriving their just powers from the consent of the governed that whenever any form of government becomes destructive of these ends it is the right of the people to alter or to abolish it and to institute new government laying its foundation on such principles and organizing its powers in such form as to seem most likely to effect their safety and happiness

a b c d e f g h i k l m n o p q r s t u x y z

Nulles ff ⌐ ⌐ d   Dowbleth σ

and for with that if but where as of the from by

so not when there this in wich is what say me my wyrt

send lr̃e receave bearer I pray you Mte your name myne

Using the One Time Pad:

1. Represent the message as a sequence of 0s and 1s of length l

2. Generate a key of length l and composed of 0s and 1s

3. XOR the message and the key

Breaking the One Time Pad:

- Ciphertext only: all the messages of same length have equal probability

- KPA, CPA, CCA: only reveal part of the key used during the attack

A *block cipher* encrypts several letters at once:

- Changing one letter in the plaintext impacts several letters in the ciphertext

- Frequency analysis of letters and digrams cannot be applied

Hill cipher:

- Invented in 1929

- One of the first cipher to use algebraic methods

- Never been used much in practice

Definition

The *greatest common divisor* of two integers a and b, with $|a| + |b| \neq 0$, is the largest positive integer dividing both a and b. It is noted $\gcd(a, b)$, and a and b are said to be *coprime* if $\gcd(a, b) = 1$.

In fact $\gcd(a, b)$ can be expressed as a linear combination of a and b with integer coefficients.

Lemma   (Bézout's identity)

Let a and b be two integers where at least one of them is not zero, and $d = \gcd(a, b)$. Then there exists two integers s and t, called *Bézout coefficients*, such that $as + bt = d$.

Algorithm. (*Extended Euclidean Algorithm*)

Input : a, b, two positive integers
Output: $r_1 = \gcd(a, b)$ and $\langle s_1, t_1 \rangle$, Bézout coefficients
1   $r_0 \leftarrow b$; $r_1 \leftarrow a$;
2   $s_0 \leftarrow 0$; $s_1 \leftarrow 1$;
3   $t_0 \leftarrow 1$; $t_1 \leftarrow 0$;
4   **while** $r_0 \neq 0$ **do**
5     $q \leftarrow r_1 \,\mathrm{div}\, r_0$;
6     $\langle r_1, r_0 \rangle \leftarrow \langle r_0, r_1 - qr_0 \rangle$;
7     $\langle s_1, s_0 \rangle \leftarrow \langle s_0, s_1 - qs_0 \rangle$;
8     $\langle t_1, t_0 \rangle \leftarrow \langle t_0, t_1 - qt_0 \rangle$;
9   **end while**
10 **return** r_1, $\langle s_1, t_1 \rangle$

Proposition

Let a and n be two coprime integers and s and t be such that $as + nt = 1$. Then $as \equiv 1$ mod n, and s is called the *multiplicative inverse* of a modulo n. Besides s is unique.

Example. What is the multiplicative inverse of 11111 modulo 12345?

Running the extended Euclidean algorithm confirms that 11111 and 12345 are coprime and therefore 11111 is invertible modulo 12345. Moreover since

$$11111 \cdot 2471 + 12345 \cdot (-2224) = 1,$$

we conclude that $11111 \cdot 2471 \equiv 1$ mod 12345.

Theorem (Cramer's rule)

Let A be an $m \times m$ matrix, then

$$\text{Adj}(A) \cdot A = \det(A)\, I_m, \tag{1.1}$$

where $\text{Adj}(A)$ denotes the adjugate of A, $\det(A)$ the determinant of A, and I_m the $m \times m$ identity matrix.

From equation (1.1) we see that for A to be invertible, $\det(A)$ must be invertible. In particular if A is defined modulo n, $\det(A)$ must be invertible modulo n, that is there exists t such that

$$\det(A) \cdot t \equiv 1 \bmod n.$$

Example. Compute the inverse of the matrix

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 4 & 9 \end{pmatrix} \mod 11.$$

Since $\det(A) = 2$ and $\gcd(2, 11) = 1$, A is invertible modulo 11 and

$$A^{-1} = \frac{1}{2} \begin{pmatrix} +\begin{vmatrix} 2 & 3 \\ 4 & 9 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 4 & 9 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 2 & 3 \end{vmatrix} \\[2mm] -\begin{vmatrix} 1 & 3 \\ 1 & 9 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 1 & 9 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & 3 \end{vmatrix} \\[2mm] +\begin{vmatrix} 1 & 2 \\ 1 & 4 \end{vmatrix} & -\begin{vmatrix} 1 & 1 \\ 1 & 4 \end{vmatrix} & +\begin{vmatrix} 1 & 1 \\ 1 & 2 \end{vmatrix} \end{pmatrix} \mod 11.$$

Then calculating all the cofactors yields

$$A^{-1} = \frac{1}{2} \begin{pmatrix} 6 & -5 & 1 \\ -6 & 8 & -2 \\ 2 & -3 & 1 \end{pmatrix} \text{ mod } 11.$$

In this case it is easy to see that 6 is the inverse of 2 modulo 11, such that we get

$$A^{-1} = \begin{pmatrix} 36 & -30 & 6 \\ -36 & 48 & -12 \\ 12 & -18 & 6 \end{pmatrix} \equiv \begin{pmatrix} 3 & 3 & 6 \\ 8 & 4 & 10 \\ 1 & 4 & 6 \end{pmatrix} \text{ mod } 11.$$

Constructing Hill cipher:

- Key: generate a random $n \times n$ matrix K modulo 26, with $\gcd(\det(K), 26) = 1$

- Encrypt:
    - Split the plaintext into blocks of size n, padding with extra letters if necessary
    - Multiply each block considered as a vector by the matrix K

- Decrypt:
    - Split the ciphertext into blocks of size n
    - Multiply each block considered as a vector by the matrix K^{-1}

Example. Encrypt "good morning" with the key $K = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 7 & 8 \end{pmatrix}$.

1. Split and pad the plaintext

g	o	o	d	m	o	r	n	i	n	g	x
6	14	14	3	12	14	17	13	8	13	6	23

$\underbrace{\qquad}_{A} \quad \underbrace{\qquad}_{B} \quad \underbrace{\qquad}_{C} \quad \underbrace{\qquad}_{D}$

2. Multiply each vector by K

$\overbrace{\qquad}^{A'} \quad \overbrace{\qquad}^{B'} \quad \overbrace{\qquad}^{C'} \quad \overbrace{\qquad}^{D'}$

6	24	6	21	8	11	11	25	11	10	9	25
G	Y	G	V	I	L	L	Z	L	K	J	Z

Knowing "goodmorningx" and "GYGVILLZLKJZ" recover the key.

1. Find n: since $n|12$, try some values until the right one is found

2. Use the three first blocks to construct the equation

$$\underbrace{\begin{pmatrix} 6 & 14 & 14 \\ 3 & 12 & 14 \\ 17 & 13 & 8 \end{pmatrix}}_{A} \cdot \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \equiv \begin{pmatrix} 6 & 24 & 6 \\ 21 & 8 & 11 \\ 11 & 25 & 11 \end{pmatrix} \bmod 26$$

3. Since A is not invertible modulo 26, try with the three last blocks

$$\underbrace{\begin{pmatrix} 3 & 12 & 14 \\ 17 & 13 & 8 \\ 13 & 6 & 23 \end{pmatrix}}_{A} \cdot \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \equiv \begin{pmatrix} 21 & 8 & 11 \\ 11 & 25 & 11 \\ 10 & 9 & 25 \end{pmatrix} \bmod 26$$

4. Since A is now invertible we calculate

$$K = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \equiv \begin{pmatrix} 3 & 12 & 14 \\ 17 & 13 & 8 \\ 13 & 6 & 23 \end{pmatrix}^{-1} \cdot \begin{pmatrix} 21 & 8 & 11 \\ 11 & 25 & 11 \\ 10 & 9 & 25 \end{pmatrix} \bmod 26$$

$$K = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \equiv \begin{pmatrix} 11 & 18 & 4 \\ 7 & 11 & 10 \\ 1 & 22 & 11 \end{pmatrix} \cdot \begin{pmatrix} 21 & 8 & 11 \\ 11 & 25 & 11 \\ 10 & 9 & 25 \end{pmatrix} \bmod 26$$

And the key is

$$K = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 9 & 7 & 8 \end{pmatrix}.$$

Remarks on Hill cipher:

- In a substitution cipher, changing one letter from the plaintext alters one letter from the ciphertext

- In Hill cipher changing one letter from the plaintext alters the whole corresponding block from the ciphertext

- Hill cipher is not vulnerable to frequency analysis attacks

- As a drawback a small error in the transmission can induce a major error in the encrypted message and the deciphered text becomes un-readable

Device information:

- Developed in Germany during the 1920s

- 1054560 ways to initialise the machine

- 100391791500 ways to interchange six pairs of letters

- Secretly broken in Poland in the 1930s

- Techniques extended by the British during World War II
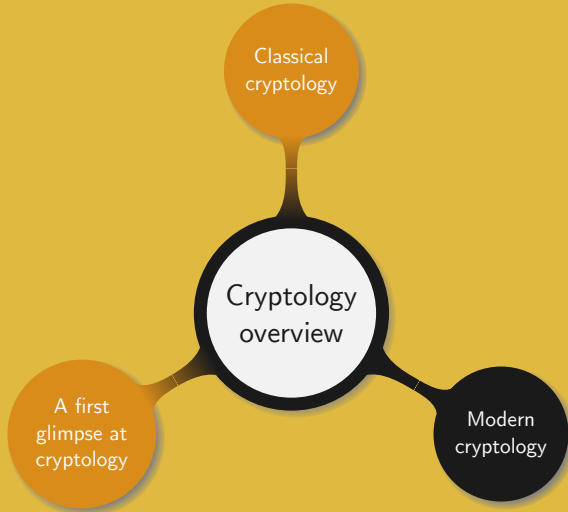
All the previous schemes are symmetric:

- The same key is used to both encrypt and decrypt

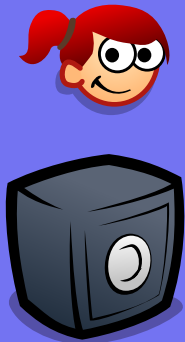- The decryption key is easily derived from the encryption key

Limitations:

- Alice and Bob need to meet in order to exchange, generate, or share the secret keys

- Key management problem:
  - 2 users $\rightarrow$ 1 key
  - 5 users $\rightarrow$ 4 keys each, total 10 keys
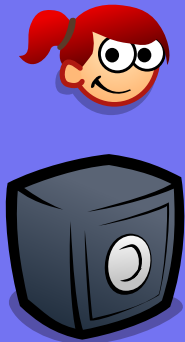 - n users $\rightarrow$ $n - 1$ keys each, total $O(n^2)$ keys

Anybody can lock the padlock but only Bob can unlock it

Mathematical problems used in Public Key Cryptography (PKC):

- Easy to generate by anybody

- Hard to solve for everybody

- Easy to solve when knowing a small secret

Common examples:

- Multiplication and factorisation

- Exponentiation and discrete logarithm problem

Over time security has depended on:

- Early years: keeping the encryption method secret

- After WW I: keeping the secret key unknown

- Modern cryptography:
  - The method, the encryption key, and how to find the secret key are known
  - Security depends on the computational infeasibility of finding it

  PKC adds much flexibility at a high computational cost

Basic security feeling:

- Obvious strategy: brute force all possible keys

- Intuition: the larger the key space the harder finding the key

Example. Substitution cipher:

- Key space: $26! \approx 4 \cdot 10^{26} \approx 2^{89}$

- Very simple to break using frequency analysis

Brute force is to be used only if no other attack is possible

Best CPUs available in 2015:

- Regular user: 298,190 MIPS (Intel Core i7 5960x)

- Supercomputer: 10,000,000,000 MIPS (Fujitsu K – 705,024 cores)

How many such computers need to run for a year to complete a program composed of 2^{80} instructions?

Very easy	Easy	Hard	Very hard	Secure
2^{56}	2^{64}	2^{80}	2^{128}	
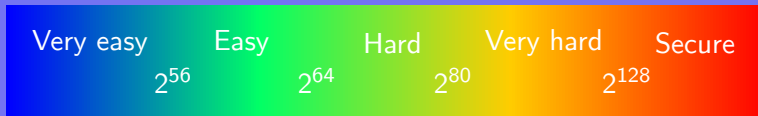
The goal is to be secure in the worst case

In the worst case the attacker:

- Has the best computational facilities

- Uses the most efficient attack available

To be secure against such an attacker:

- Check to complexity of the best algorithm available

- Adjust the parameters of the cipher such that more than 2^{128} operations are required to break the encryption

Example. Assuming that the best attack on a mathematical problem requires $\sqrt{n}$ operations, where n is the size of the key, what key size should be chosen to be secure?

Since secure means that the attacker has to compute at least 2^{128} operations to break the encryption it suffices to calculate

$$\left(2^{128}\right)^2 = 2^{256}.$$

Hence the key space should contain 2^{256} elements, that is the key should be at least 256 bits long.

Is double encryption with two different keys enhancing security?

Improving security:

- Naive answer: for a key of length k, 2^{2k} operations are needed

- Better answer:
    - It does not change anything, e.g. Hill cipher
    - It is possible to do better than 2^{2k}: meet in the middle attack

Symmetric encryption using a function f and a key k:

- Simple encryption: $c = f_k(m)$

- Double encryption: $c = f_{k_2}(f_{k_1}(m))$

- Decryption: $m = f_{k_1}^{-1}(f_{k_2}^{-1}(c))$

Assuming a KPA setup:

1. For all the keys, compute and store the ciphertexts $c_i = f_{k_i}(m)$

2. Compute all plaintexts $m_i = f_{k_i}^{-1}(c)$ and find any matching c_i

3. Recover the corresponding keys k_1 and k_2

4. Test k_1 and k_2 on more plaintext/ciphertext pairs

Exercise. Assuming no attack applies on an encryption scheme and a key size of 64 bits, what is its security if applying double encryption?

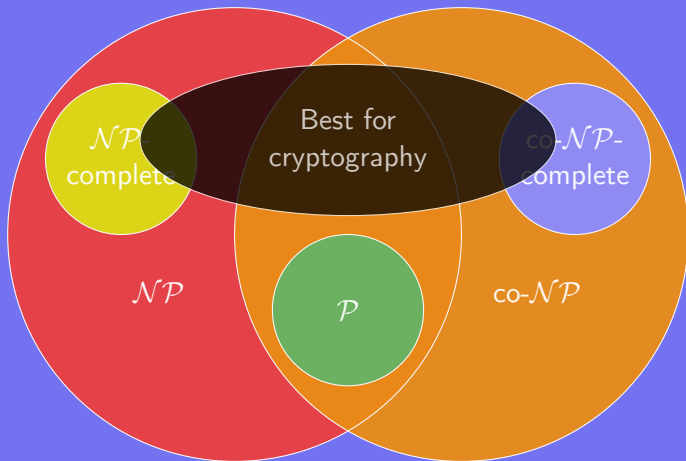Main complexity classes related to cryptology:

- $\mathcal{P}$: decision problems for which there exists a deterministic polynomial time algorithm

- $\mathcal{NP}$: decision problems for which the answer "yes" can be verified using a deterministic polynomial time algorithm

- $\mathcal{NP}$-complete: hardest problems in $\mathcal{NP}$

- co-$\mathcal{NP}$: decision problems for which the answer "no" can be verified using a deterministic polynomial time algorithm

- co-$\mathcal{NP}$-complete: hardest problems in co-$\mathcal{NP}$

Example. Integer factorization is in both $\mathcal{NP}$ and co-$\mathcal{NP}$

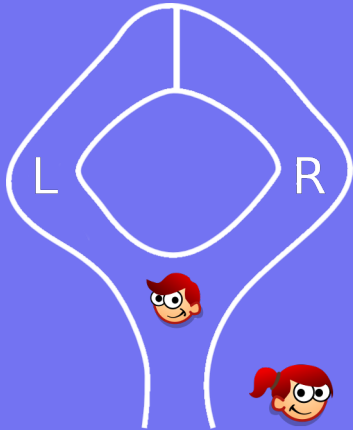Let n be a large integer and $1 < m < n$. Does n have a factor p, with $1 < p < m$?

- $\mathcal{NP}$: with certificate "p a factor of n" verify in polynomial time that $1 < p < m$ and $p|n$

- co-$\mathcal{NP}$: with certificate "the list of all the prime factors of n" verify in polynomial time that:

  - They are all prime

  - Their product is n

  - None of them is between 1 and m
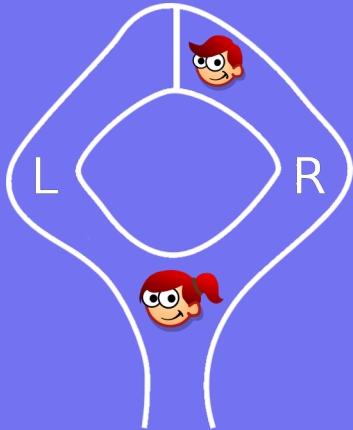
Bob knows a secret path, and wants to prove it without revealing it

Strategy:

1. Alice hides while Bob chooses to go Left (L) or Right (R)

2. Alice randomly asks Bob to exit on L or R

3. If Bob is on the wrong side he uses the secret path or otherwise returns

4. Repeat steps 1 to 3 many times

Bob knows a secret path, and wants to prove it without revealing it

Strategy:

1. Alice hides while Bob chooses to go Left (L) or Right (R)

2. Alice randomly asks Bob to exit on L or R

3. If Bob is on the wrong side he uses the secret path or otherwise returns
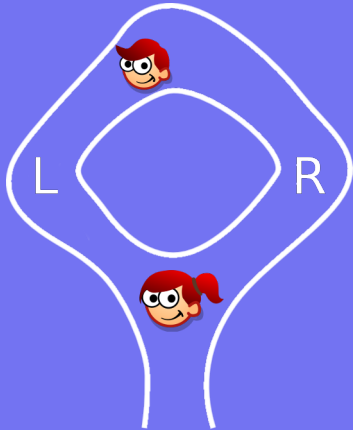
4. Repeat steps 1 to 3 many times

Bob knows a secret path, and wants to prove it without revealing it

Strategy:

1. Alice hides while Bob chooses to go Left (L) or Right (R)

2. Alice randomly asks Bob to exit on L or R

3. If Bob is on the wrong side he uses the secret path or otherwise returns

4. Repeat steps 1 to 3 many times

Definitions

① Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two simple graphs. Then we say that G_1 and G_2 are *isomorphic* if there exists a bijective function $\varphi \colon V_1 \to V_2$ such that the induced map

$$\varphi_* \colon E_1 \to E_2, \qquad (a, b) \mapsto (\varphi(a), \varphi(b))$$

is bijective. Such a function φ is called a *graph isomorphism*.

② A *Hamilton circuit* in a graph G is a simple circuit that passes through every vertex of G exactly once.

Hard problems related to graph theory:

- Graph isomorphism:

  - No known polynomial time algorithm

  - Not proven to be $\mathcal{NP}$-complete

  - Best known algorithm has exponential complexity

- Finding a Hamiltonian circuit:

  - Proven to be $\mathcal{NP}$-complete

  - Best known algorithm has exponential complexity

Initial setup:

- A graph G
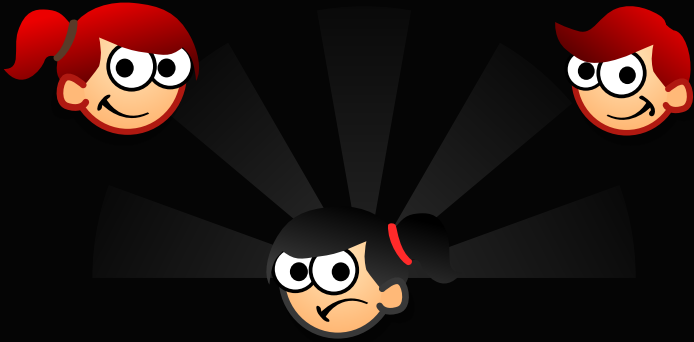- A Hamiltonian circuit in G

- Bob's graph G

Process:

1. Bob generates H, a graph isomorphic to G

2. Bob commits H

3. Alice randomly asks for either the isomorphism or a Hamiltonian circuit in H

4. Bob either shows the isomorphism or translates the Hamiltonian circuit in G onto H and shows it

Regarding the process and setup:

- Is Bob revealing any sensitive information?

- Why does Bob need to commit H?

- If Bob can anticipate Alice's request how can he cheat if asked for:
  - The graph isomorphism?
  - The Hamiltonian circuit?

- Explain whether randomness is important or not in the process.

- Can any of Bob and Alice cheat?

- How many times should the process be repeated for Bob to prove that he really knows a Hamiltonian circuit in G?

- What is cryptology?

- Who are Alice, Bob, and Eve?

- What is Kerckhoff's principle?

- Explain the One-Time-Pad

- Explain the underlying idea of public key cryptography

- In 2022 what security level is considered safe?

Thank you!

1.12 https://www.xkcd.com/538/

1.28 Simon Singh, *The Code Book – How to make it, break it, hack it, crack it?*

1.42 https://cdn.globalauctionplatform.com/7187abcf-14de-4d26-9a48-a48e012a3b
d3/1f194fa9-87f4-45cd-b6db-a48e012de5c7/original.jpg