

# Foundations of Information Processing

## Data encryption

# Considerations about data encryption

## Consideration 1:

How to encode data  
to be **easy** to read and to interpret ?

## Consideration 2:

How to encode data  
to be difficult to read and to interpret?

## Consideration 3:

Is there the optimal way  
to encrypt data  
from outsiders?

Data can be encoded again in such a way that  
the representation of data  
changes to a different form.

If the reader does not know  
the way of encoding  
it can be very difficult to read data.

# Information security and data encryption

- Concepts: steganography, cryptography, cryptology.
- Symmetric-key encryption: the same key for encryption and for decryption.
- Public-key encryption (asymmetric): the public key for encryption and the private key for decryption.
- Examples about cipher machines and cryptosystems: Enigma and Lorenz, Vernam and RSA.
- Sources for further information:
  - Brookshear, J.G. *Computer Science - An overview*, 13th ed. Addison Wesley, 2019.
  - Schneier, B. *Applied cryptography: protocols, algorithms, and source code in C*, 2<sup>nd</sup> ed., Wiley, 1996.
  - Ciphers: <http://www.cryptomuseum.com/crypto/index.htm>
  - Bletchley Park: [YouTube](#), for example, [Bletchley Park Tour](#).

# Steganography

- Concealing a message within another message or a physical object.
- Ancient Greek:
  - "steganos" = "covered or concealed".
  - "graphein" = "to write".
- Examples:
- "Hair encryption" in Egypt:
  - Write a message on a bald head, let the hair to grow, deliver the message to the destination physically.
- Digital signatures in digital images:
  - Hide the signature in the image using the least significant bits for the verification of the originality.
  - The signature is invisible to the human, but can be decoded.
- Hidden meaning of text:
  - For example, the first letters matter, "Still another idea put away"  
=> what is the lecturer's favorite ice-hockey team?

# Cryptography and cryptology

- **Cryptography:**

- The scientific field to secure communication, keeping messages hidden.
- "kryptós" = "hidden, secret"
- "graphein" = "to write".
- "cipher" (engl.) = a secret way of writing, especially one in which a set of letters or symbols is used to represent other.

- **Cryptology:**

- The science concerned with data communication and storage in a secure and usually secret form.
- "logia" = studies.
- A message is encrypted from third parties.
- Methods as principles (cryptosystems), devices as help (cipher machines) .
- For example: Enigma and Lorenz.



# Encryption and decryption

- Encryption is the process to **encode** the original information of the message (**plaintext**) to the hidden representation (**ciphertext**) in order to secure information.
- **Encryption** (encoding) by the function  $E$ :
  - From the original plaintext (message)  $M$  to the ciphertext  $C$ :

$$E(M) = C$$

- **Decryption** (decoding) by the function  $D$ :
  - From the ciphertext to the original plaintext:

$$D(C) = M$$

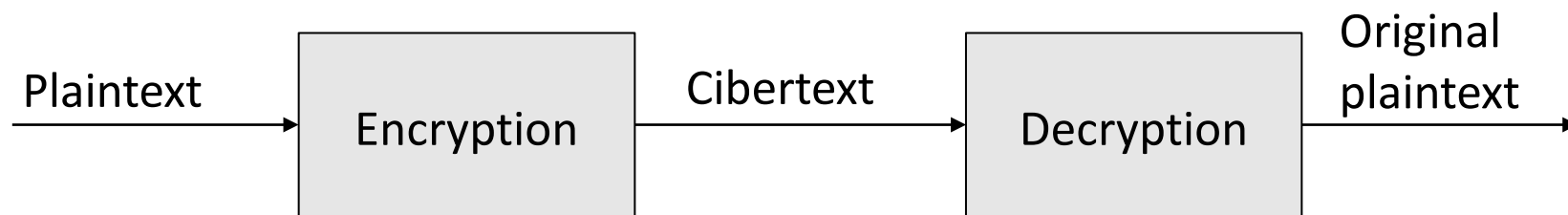
$$\Rightarrow D(E(M)) = M$$

# Information security: concepts

- Key:
  - A tool produced for encryption or a signature that converts a general encryption algorithm to a specified encryption method.
- Public key:
  - The key which is publicly available for encryption or a signature.
- Cryptanalysis:
  - The process of the analyzing information systems in order to understand hidden aspects of the systems.
  - Conclude the plaintext or the key without the information about the key.
- Concepts of cryptography:
  - **Confidentiality**: the cryptosystem is known to the enemy, expect the key to decrypt is not public, and the enemy can freely experiment with it (the Kerchoff's principle).
  - **Integrity**: the content of the message does not change.
  - **Authenticated** and **undeniable** data: the message is authenticated by the public key.

# Secret algorithms

- The encryption of messages started using secret algorithms.
  - This approach became common after direct steganography where the message was hidden, but it was not yet encoded/decoded by an algorithm.
- Only the communicating parties should know the encryption method.
  - ⇒ the method should be changed if the method was revealed.



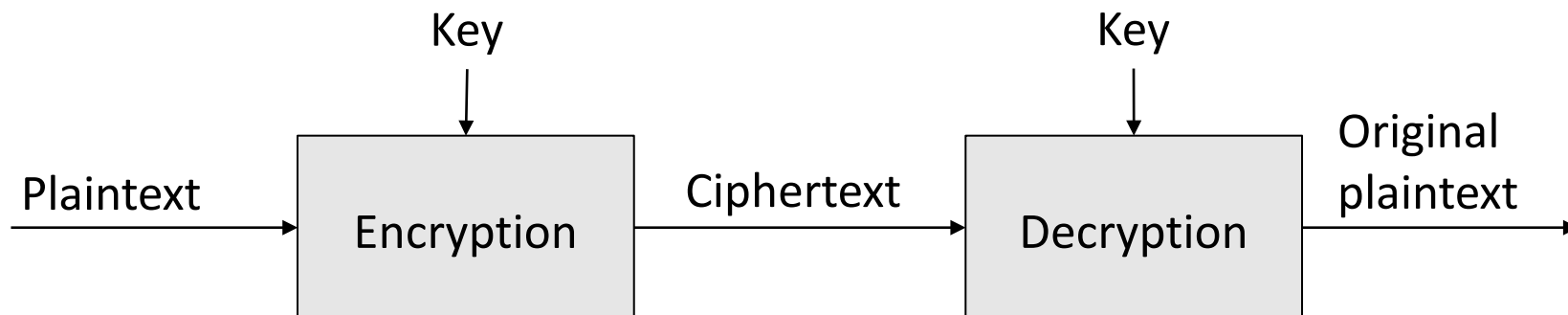
# Symmetric-key encryption

- Messages are encrypted by a known or secret algorithm using one key  $K$  as extra information:

$$E_K(M) = C$$

$$D_K(C) = M$$

- If the key is revealed, only the key is changed.
- If some weakness is found in the method, the method has to be changed.



# Asymmetric-key encryption

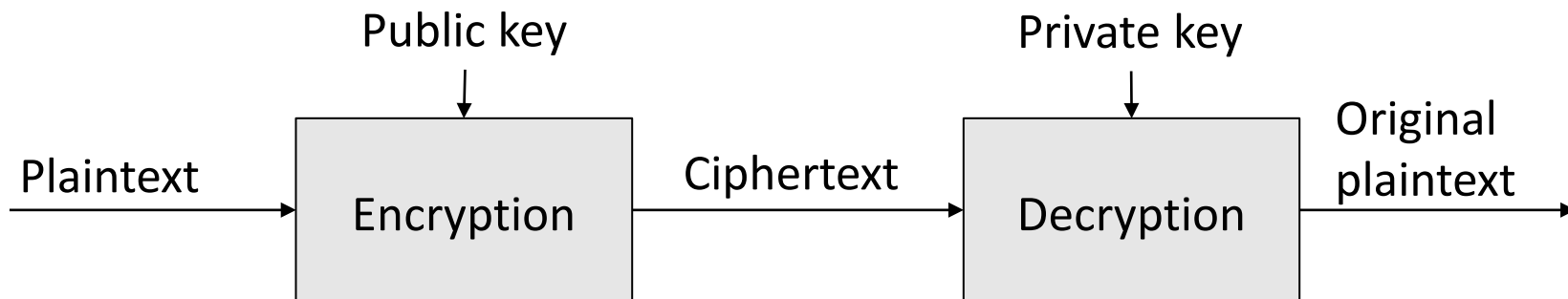
- Messages are encrypted by a known or secret algorithm using two keys, the one for encryption  $K_1$  (the public key) and the other for decryption  $K_2$  (the private key):

$$E_{K_1}(M) = C$$

$$D_{K_2}(C) = M$$

$$\Rightarrow D_{K_2}(E_{K_1}(M)) = M$$

- Methods:
  - Stream ciphers (bit by bit, or byte by byte).
  - Block ciphers (data block by data block).



# Encryption methods: substitution cipher

- **Substitution cipher** is a method of encryption where **symbols** of the plaintext (usually characters) are changed to **other symbols** according a selected strategy, but their **position** in the plaintext remains **same**.
- **Simple substitution:**
  - The characters of a message are replaced by the characters from the same alphabet.
  - Caesar (introduced by): the alphabet is shifted by 3 places forward (e.g., the character “a” becomes “d”) which makes the content to disappear (unreadable).
  - ROT13: "rotate by 13 places", the English alphabet.
    - For example, for hiding suspicious internet discussions from direct visibility.
- **Homophonic substitution:**
  - Many options to substitute a character.
  - In the replacement alphabet there are one or more options.
  - For example, for the letter E there four options Z, 7, 2 or 1 are set. A reasonable practice: the more common the letter, the more options.

# Encryption methods: substitution cipher (cont.)

- **Polygram substitution:**
  - All the characters of a string are substituted by the set of other character representations.
  - For example, the Romans wrote messages in Latin using the Greek alphabet, instead of the Latin alphabet (introduced by Ceasar).
  - Huffman encoding.
- **Polyalphabetic substitution:**
  - The combination of many one-character ciphers where in which each plaintext letter is assigned more than one substitute.
  - For example, used in the American civil war.
  - Alberti cipher disks, Leon Baptista Alberti 1467.
    - Two concentric disks, attached by a common pin, which could rotate one with respect to the other.



# Encryption methods: transposition ciphers

- **Transposition cipher** is method of encryption where the **positions** held by units of plaintext (commonly characters or groups of characters) **are shifted** according to a regular system, so that the ciphertext constitutes a permutation of the plaintext.
- The positions of characters of a message are changed to others.
- **Example:** the ADFGVX cipher in WW1 introduced by a German lieutenant Fritz Nebel.
- The characters are encoded from the letters A, D, F, G, V, and X using a 2-D transposition table. AA = b, AF = a, etc., i and j share the code GD.
- **Rotor-based cipher machines:**
  - Multistep (~rotor) character scrambling mechanisms.
  - Examples: the German **Enigma** before and during WW2 (“enigma” in Greek = “a riddle”) and the German **Lorenz** during and after WW2 (built by the company C. Lorenz AG).

	A	D	F	G	X
A	b	T	a	l	p
D	d	h	o	z	k
F	q	f	v	s	n
G	g	i/j	c	u	x
X	m	r	e	w	y



# Rotor-based cypher: Enigma

- Arthur Scherbius, a German engineer, 1918.
- First, for commercial use to secure business communication.
  - Only later for military use, especially in WW2.
- Combination of electrical and mechanical parts:
  - Rotating discs (rotors):
    - 26-character alphabet per disc.
    - 3 in use of 3-5 discs.
  - The keyboard: 26 characters.
  - The plugboard: 10 cables connect the discs and the keyboard.
- Enigma scrambles the 26 letters of the alphabet.
  1. Every time when the key is pushed (a plaintext character),
  2. the corresponding cyphertext character appears defined by the discs and the plugboard, and then
  3. the new encoding is generated.
- See for further information (YouTube): [Bletchley Park Tour](#)



# Enigma: strengths and weaknesses



- 158 million million million options =  $158 \cdot 10^{18}!!$ 
  - The probability to win in the Finnish lottery is around one of  $15 \cdot 10^6$  options.
  - The discs alone generate  $151 \cdot 10^{12}$  options (3 discs of 5 selected and there are 26 characters/disc).
- Enigma's "weaknesses" in WW2:
  - The encryption solution never generated the same character: e.g., "a" is never encoded as "a".
  - Each message contained "Heil Hitler" in the same certain place.
  - In all submarines (U-boats) there were codebooks for initiating an Enigma machine which contained secret information (and could be lost to the enemy).
- **Alan Turing**, an English mathematician and computer scientist (University of Cambridge, Princeton Univ) led the operation, called Ultra.
  - The Ultra group broke the code of Enigma by using **a computing machine** which was developed further as a programmable machine which is nowadays called as **the computer**.

# Enigma => Lorenz and Bombe => the computer

1940: Ultra Operation (HUT8, Bletchley Park, UK)  
built the Bombe computing machine.

- The encryption of Enigma was broken.

1941: the Germans built the Lorenz cipher.

- 12 cipher discs! =>  $1.6034 \cdot 10^{19}$  options.
- Enigma: "only"  $1.51 \cdot 10^{14}$  options.

1943: Colossus (two last images in the right).

- **The world's first programmable computer.**
- Designed by **Alan Turing** (planned on paper/in his mind already much earlier) and built by mathematician **Bill Tutte** and engineer **Tommy Flowers**.
- Tutte and Flowers gained recognition with a significant delay since the work was top secret.
- Turing committed suicide as a victim of forced chemical castration due to his homosexuality in 1954 (the official apology by the British government in 2009).

The British broke the encryption of Lorenz, but kept it secret during the cold war.

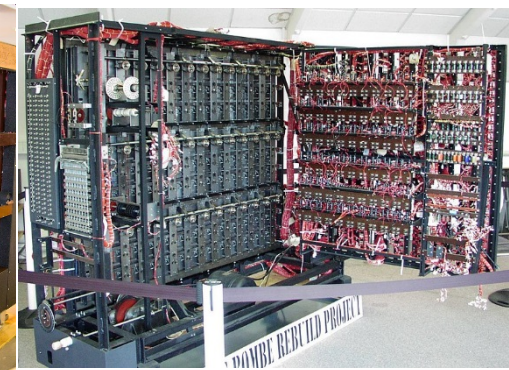
Enigma



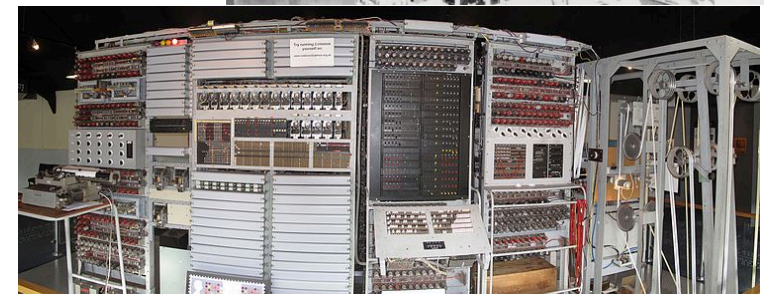
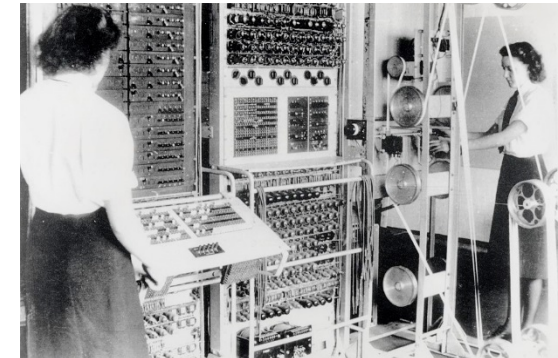
Lorenz



Bombe



Colossus



# Ciphers: more information

- Ciphers:

<http://www.cryptomuseum.com/crypto/index.htm>

- Bletchley Park:

- YouTube:

[http://www.youtube.com/results?search\\_query=bletchley+park](http://www.youtube.com/results?search_query=bletchley+park)

- For example, [Bletchley Park Tour](#).



# Considerations about encryption methods

- Depending on a method, if either
  - the encryption method itself or
  - the used/required key is revealedthe decoding of the plaintext is much easier.
- Is there the optimal way to secure the plaintext?  
=> the single-use pre-shared key!
- One-time pad (OTP): plaintext is paired with a random secret key (called the one-time pad). Then, each bit or character of the plaintext is encrypted by combining it with the corresponding bit or character from the pad using modular addition.
- Impossible to decrypt or break if the following requirements are met:
  - The key is random, is at least as long as the plaintext, is used once only, and is kept secret by the communicating parties.
- Is it too slow and impractical for real-time communication?

# Vernam Cipher: one time pad

- Vernam Cipher, Gilbert S. Verman, Bell Labs, U.S. Patent, 1919 to encrypt teletype messages (proposed in 1917).
- The Vernam Cipher combines the plaintext (the original message) with pseudo-random series of polyalphabetic characters to form the ciphertext using the “exclusive or” (XOR) function.
- The XOR operation for bits is defined as:

$$0 \oplus 0 = 0$$

$$0 \oplus 1 = 1$$

$$1 \oplus 0 = 1$$

$$1 \oplus 1 = 0$$

- The following must hold too:

$$a \oplus a = 0$$

$$a \oplus b \oplus b = a$$

# How to use XOR?

- The method (M = the message, K = the key, C = the cyphertext):

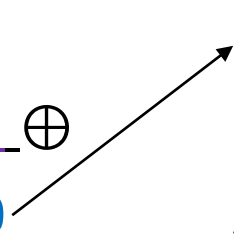
$$M \oplus K = C$$

$$C \oplus K = M$$

- Encode M with XOR when the key is K, and decode using K back to M.
  - plaintext** + **key** = **ciphertext**  $\Rightarrow$  **ciphertext** + **key** = **plaintext**

- Example:

Plaintext	A	00011		G	11010
Key	B	<u>11001</u>	$\oplus$	B	<u>11001</u> $\oplus$
Ciphertext	G	11010		A	00011



The characters represented by bits are from a codebook.

# One-time pad

- The key is used *once only*:  
⇒ No use of finding the key by cryptanalysis since it is not used again.
- If the key is *as long as* the plaintext itself:  
⇒ cryptanalysis is not applicable since all messages of plaintext are equally as probable.

⇒ The perfect encryption method:

$M = \text{ONETIMEPAD}$

$K = \text{TBFRGFARFM}$

$C = \text{IPKLPSFHGQ}$

$(O+T \bmod 26 = I$

$N+B \bmod 26 = P$

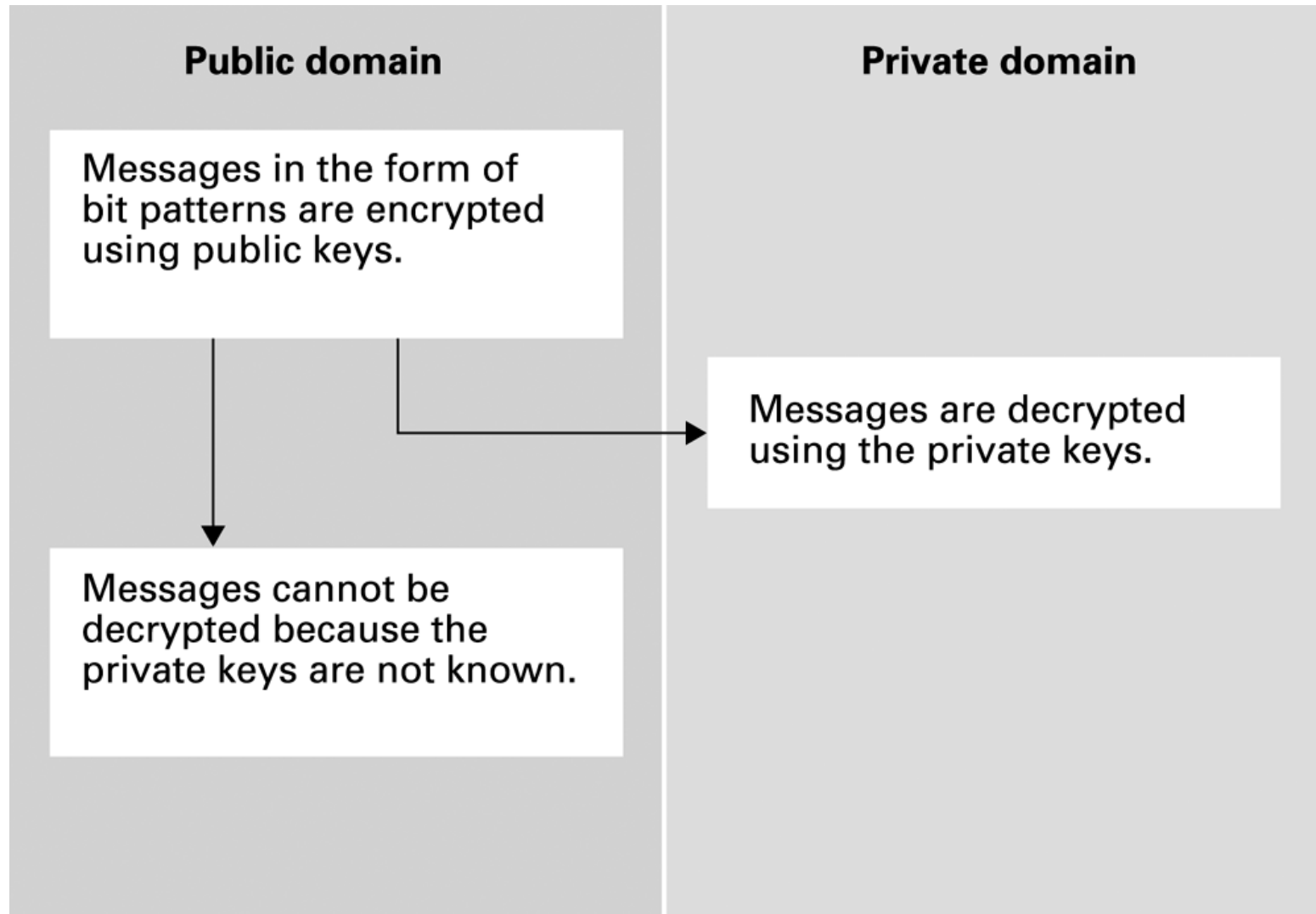
...)



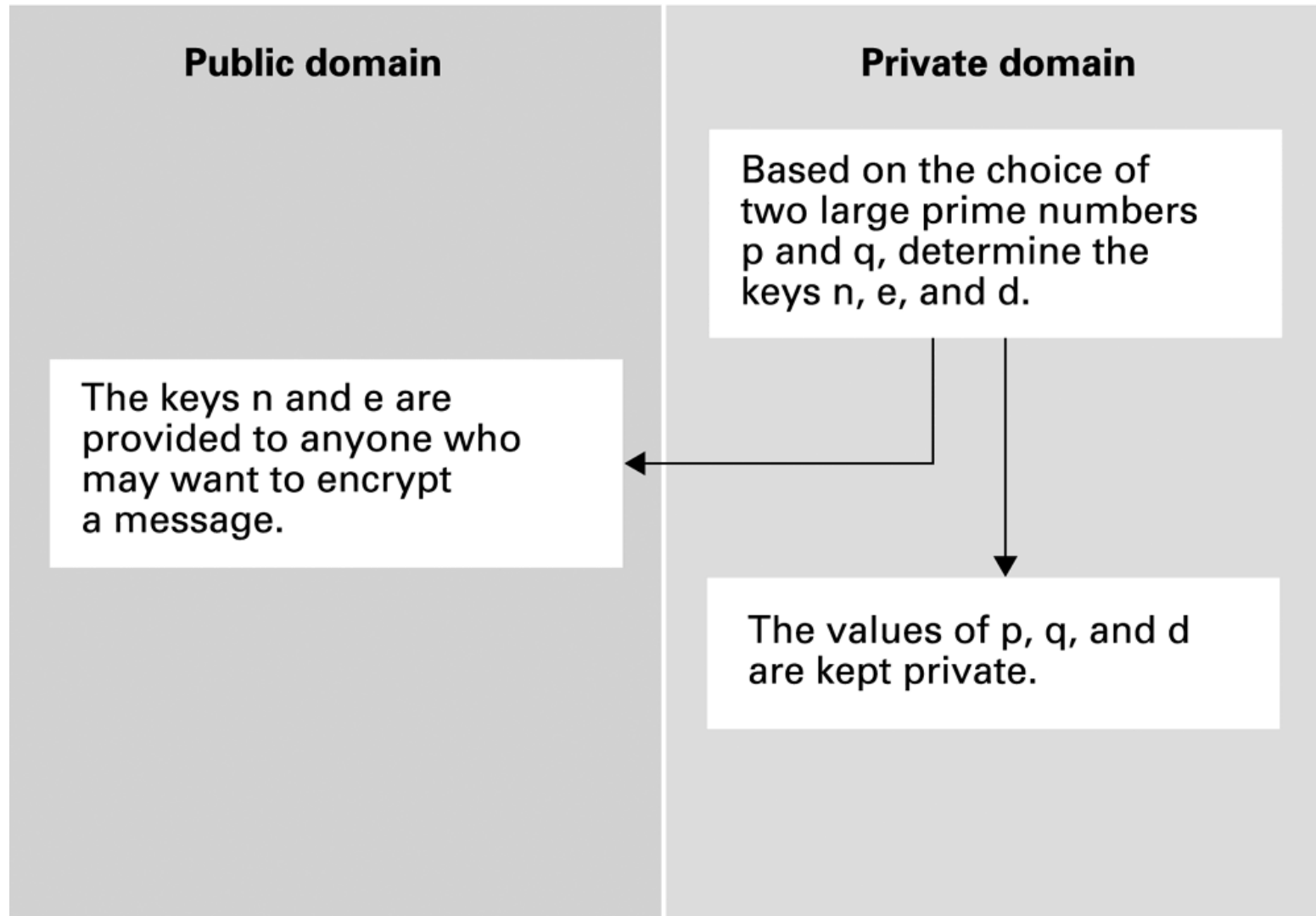
# Asymmetric-key encryption

- The solution how to distribute the keys: public key methods.
- The public key: a tool which is used to encrypt messages (or to authenticate signatures).
- The secret key: a tool which is used to decrypt secure messages (or to generate signatures).
- Example: the sender of a message has encoded the message as  $100_2$  using the public key given by the receiver, and for decoding of the message the receiver uses his/her secret key which the others do not know.
- RSA cryptosystem:
  - Rivest (MIT), Shami (Tel Aviv, Weizmann), Adleman (UCB), 1978.
  - Equivalently invented by Clifford Cocks, Government Communications Headquarters (GCHQ), 1973 (confidentially secret till 1997).
  - A very popular encryption method which uses the public keys.
  - The public keys are based on two large prime numbers and their expected complexity in calculations.
  - Easy to calculate forward (to encrypt), but very difficult to calculate backwards (decrypt), if all the keys are not known.

# Public key encryption



# RSA cryptosystem



# Example: the RSA keys

- Generating the keys:

$$n = pq$$

$$ed = k(p-1)(q-1)+1$$

Encryption:  $n$  ja  $e$  are the public keys

Decryption:  $n$  ja  $d$  ( $d$  is the secret key)

- For example:

$$p = 7$$

$$q = 13$$

$$n = pq = 7 \cdot 13 = \mathbf{91}$$

$$\mathbf{e = 5}$$

$$\mathbf{d = 29}$$
 since

$$ed = k(p-1)(q-1)+1$$

$$\Leftrightarrow 5 \cdot 29 = 2(7-1)(13-1)+1$$

$k=1$  would generate

$$1(7-1)(13-1)+1=73$$

which is a prime

so it could not be divided.

$$2(7-1)(13-1)+1=145$$

is not a prime.

# RSA: encryption

- Example: the encryption of the message

$$M = 10111_2$$

- Keys for encryption:  $n = 91$ ,  $e = 5$

$$10111_2 = 23_{10}$$

$$23^e = 23^5 = 6\,436\,343$$

$6\,436\,343 / 91$  produces the remainder 4

$$4_{10} = 100_2$$

- Thus, the message  $M = 10111_2$  is encrypted as follows:

$$C = 100_2$$

# RSA: decryption

- Decryption of the ciphertext message:

$$C = 100_2$$

- Keys for decryption:  $d = 29$ ,  $n = 91$

$$100_2 = 4_{10}$$

$$4^d = 4^{29} = 288\ 230\ 376\ 151\ 711\ 744$$

$288\ 230\ 376\ 151\ 711\ 744 / 91$  produces the remainder 23

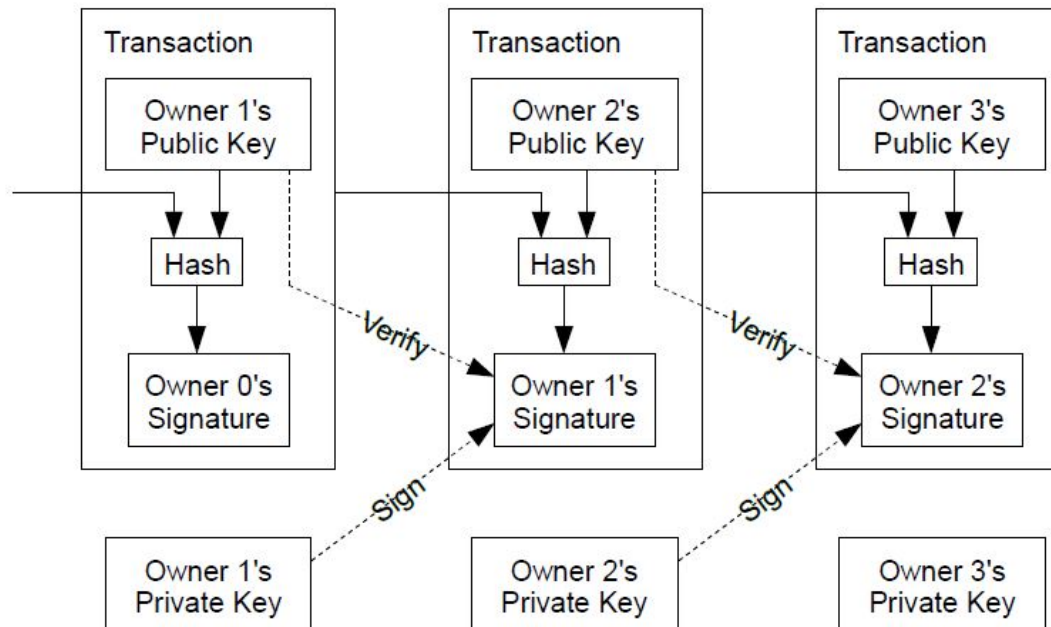
$$23_{10} = 10111_2$$

- Thus, the ciphertext  $C = 100_2$  is decrypted as follows:

$$M = 10111_2.$$

# Blockchain

- A blockchain is a growing list of records, called blocks, that are linked together using cryptography.
- Each block contains a cryptographic hash of the previous block, a timestamp, and transaction data.
- Data is authenticated besides securing it, and it is “impossible” to forge.
- Satoshi Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System, Decentralized Business Review, 2008.



# Cryptoanalysis and cyberattacks

- In cryptoanalysis,  
the plaintext and the key are tried to be deciphered  
without (exact verified) information  
about the encryption method or the key used.
- Methods to decipher the key or the encryption algorithm:
  - Microscopy: data is “hidden”, micro marks, micro points (for example, printers and passports).
  - Linguistics.
  - Frequency analysis, combinatorics, probabilities.
  - Attacks:
    - Different ways to utilize plaintexts, ciphertexts, keys, and guesses about them.
- Cyberattacks are the challenge in a larger scale:
  - A cyberattack is any offensive maneuver that targets computer information systems, computer networks, infrastructures, or personal computer devices.



# Summary

- In many information processing applications, the data is not for everyone to see due to **confidentiality**, and in some cases electronic documents require **authenticated** and undeniable signatures.
- Information security provides important tools for encoding and decoding of the data safely.
- Cipher machines: Enigma and Lorenz.
- Data encryption is based on symmetric algorithms or **asymmetric algorithms** (the public and private keys).
- The perfect encryption method exists: the one-time pad.