Information Coding
2013/2014

Lecture 11

18 November 2013

Summary:

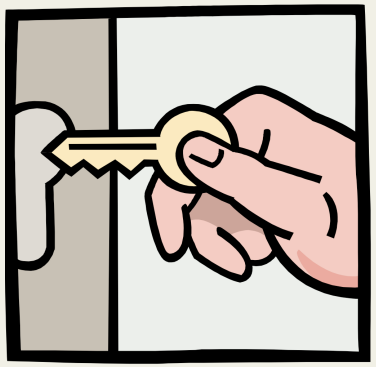
Cryptography

- Introduction, terminology, classification criterions
- Some classical cipher methods
- Perfect secret (Shannon)
- Steganography

Cryptography and Crypto-analyses

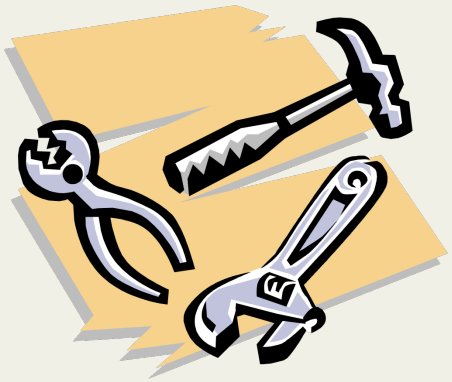
Cryptography

The art of making codes

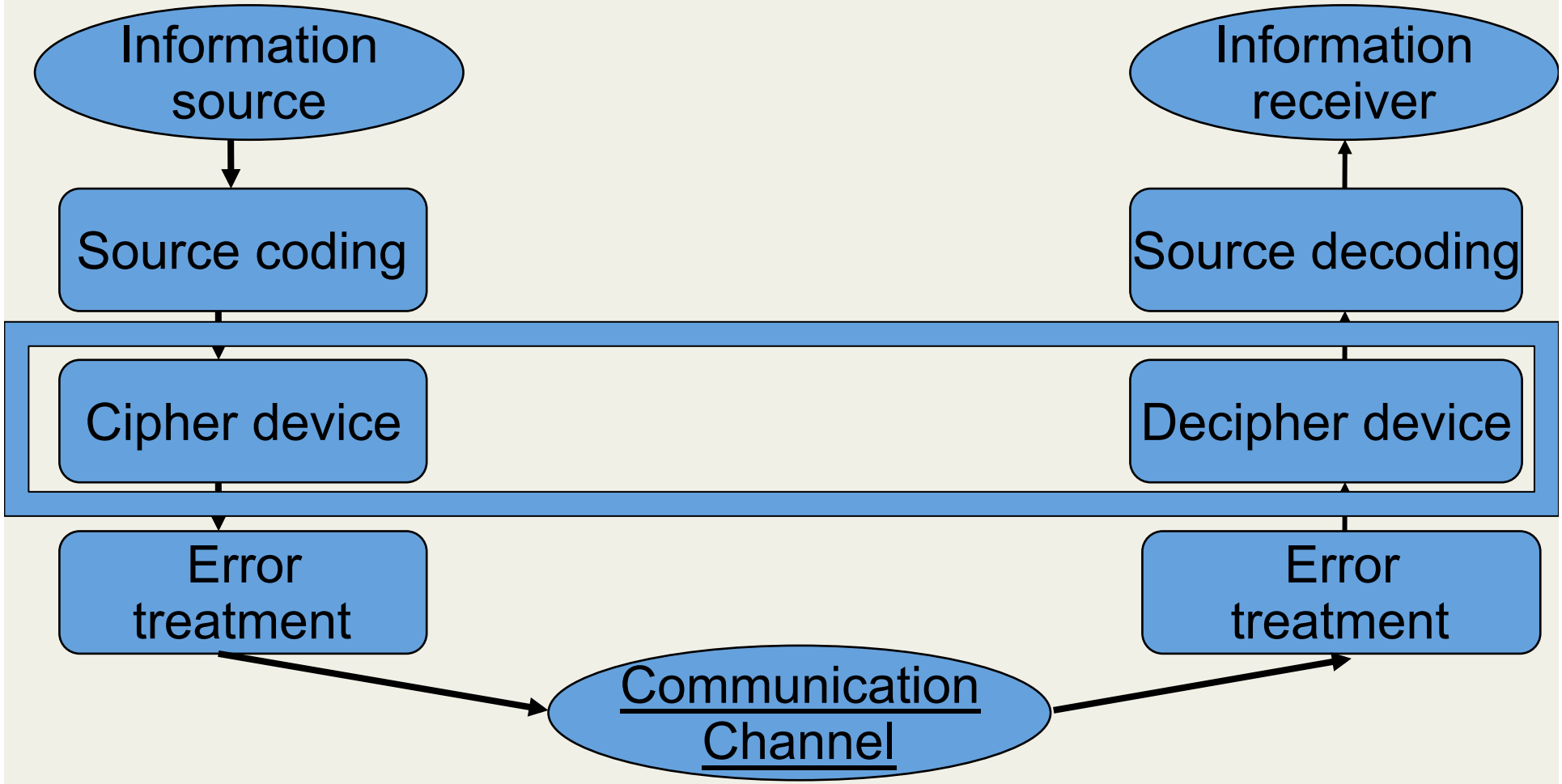


Crypto-analyses

The art of breaking codes



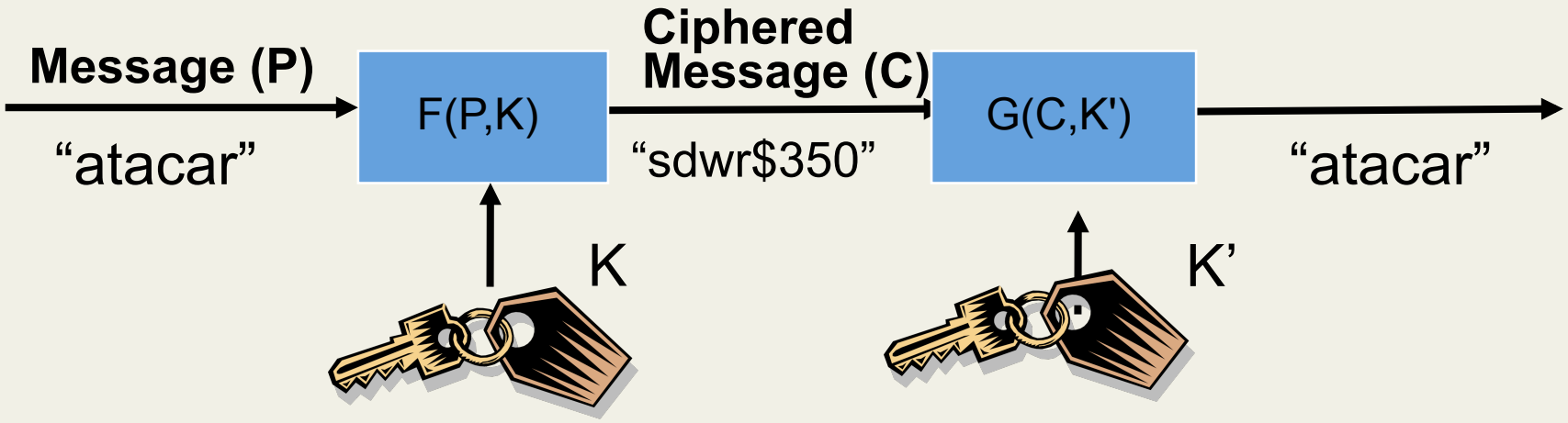
Cryptography



Cipher or Cipher system

A cipher or cipher system is basically a pair of functions:

- $\mathbf{F_K}$ (cipher function) makes the correspondence between a set P and a set C; $\mathbf{F_K}$ is based on a value **K** that is known as the **cipher key**;
- $\mathbf{G_{K'}}$ (decipher function) is the inverse function of $\mathbf{F_K}$; $\mathbf{G_{K'}}$ is based on a value **K'** that is known as the decipher **key**.



Symmetric and asymmetric cryptography

Based on C determining P without knowing K' is very difficult

- If K = K´ (or is easy to determine K' based on K), we have **symmetric cryptography** (**secret key cipher**)
- If is extremely difficult to calculate K' based on K, we have **asymmetric cryptography**:
  - On the cipher systems that use this approach, many times K is known by all (**public key**) and K' is only known by the message receiver (**private key**)
  - The public key cryptography is relatively recent (1976)

Crypto-analyses

Attacks with ciphered text only

The attacker has only two ciphers (with the same algorithm) of several messages. The goal is to recover the original text, or better, find out ("deduce") the original key or keys.

Attacks with a known simple text

The attacker has some ciphers (with the same algorithm) of known messages. The goal is identical to the previous one.

Attacks with a chosen simple text

Similar to the previous one, but in this one is the cryptanalyst who chose the text to be ciphered.

Crypto-analyses

Adaptive attack with a simple text chosen

This a special case of the previous one where the texts to send don't need to be all known in advance, they can be chosen after the results of the previous coding is known

Attack with a chosen ciphered text

Cryptanalysts can choose different ciphered texts and have access to the texts after the decoding process.

Ciphers and Codes

Ciphers

- Transform fixed size units, by using a function that describes who to code a unit
- Not depended of the language used

Codes

- Dictionary that makes a correspondence between words
- Are language dependent so their coding depend on the language used
- A code book has the dictionary
- Is difficult to distribute, maintain and protect the code book

Cypher types

Type of key

- Symmetric ciphers
- Asymmetric ciphers
- Hybrid or mix ciphers

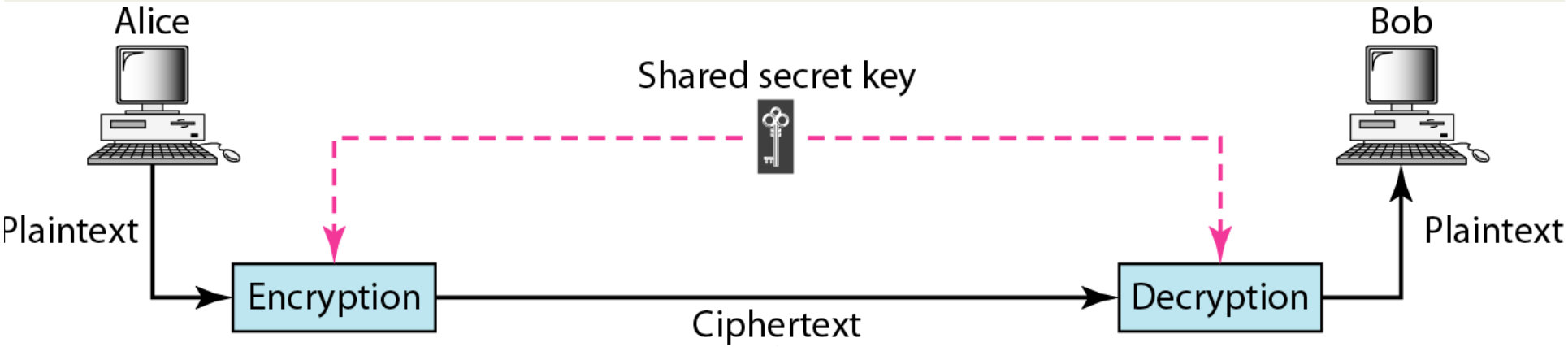
Operation modes

- Block cipher
- Stream cipher – *continuous or sequence*

Keys - symmetric ciphers

- Key common to the cipher and decipher operation
- P – original message
- C – ciphered message
- K – secret key used on the cipher and decipher operations
- f – function used to cipher
- f_{-1} – function used to decipher

$$C = f(K, P) \qquad P = f_{-1}(K, C)$$



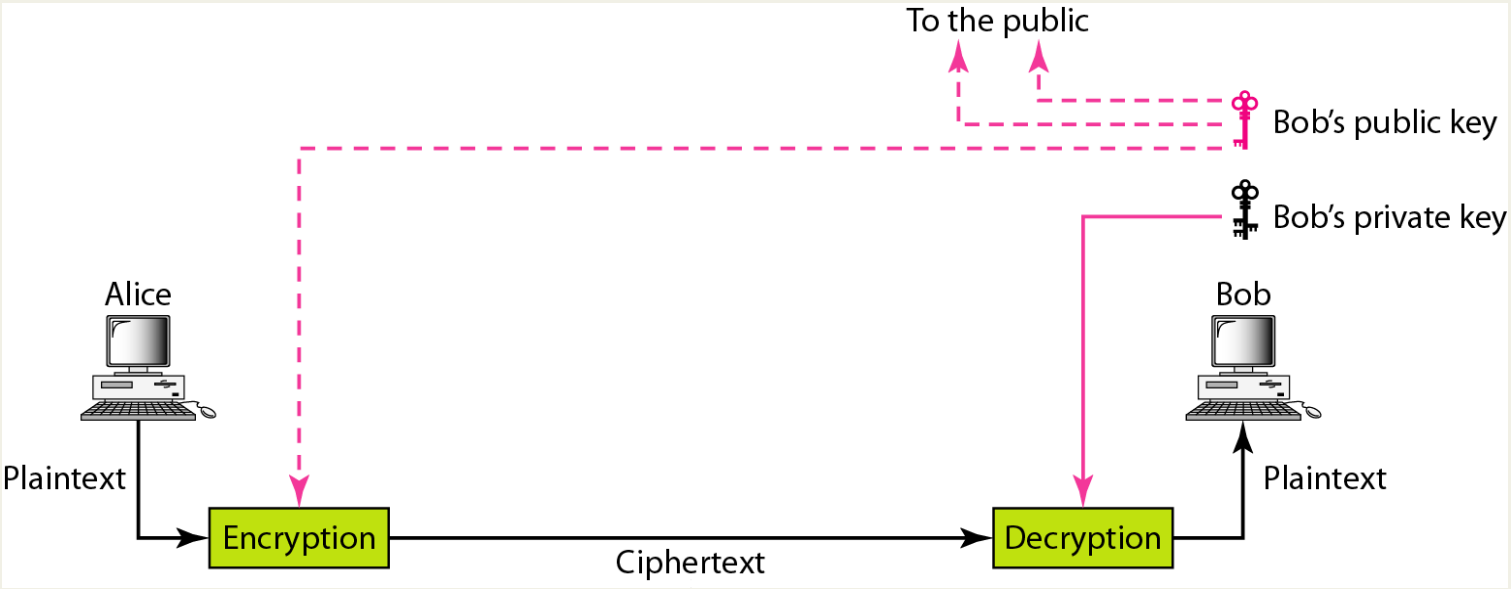
Keys - asymmetric ciphers

- K_p – Public key used to cipher
- K_s – Private key used to decipher (different from K_p)

These keys are connected to a certain entity to whom we want to communicate with

- $C = f(K_p, P)$ and $P = f_{-1}(K_s, C)$

The functions f and f_{-1} have greater execution time than the functions used with the symmetric ciphers.



Keys - hybrid

- The transfer of large amounts of data is made using a symmetric method with secret key **K**

- The key **K** exchange, between interlocutors, uses an asymmetric method (public key, private key)

Some types of classical ciphers

- Substitution ciphers
  - Monoalphabetic
  - Polialphabetic
    - Vigenère cipher
    - Vernam cipher (*one time pad*)
- Permutation cipher

Caesar cipher

Replace each character in the original text by the character 3 places ahead (key=3)

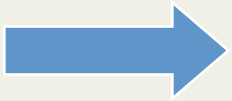
Original text	A	B	C	D	E	F	G	H	…
Encrypted text	D	E	F	G	H	I	J	K	…

A=1, B=2, C=3, …

cipher: C = P+3
decipher: P = C-3

Caesar cipher

- Monoalphabetic cipher with shift 3
- A letter in original text always corresponds to another letter in the ciphered text
- The correspondence is maintained throughout the text
- Are easily broken by statistical analysis
- T(a) = (P(a)+3) mod 26

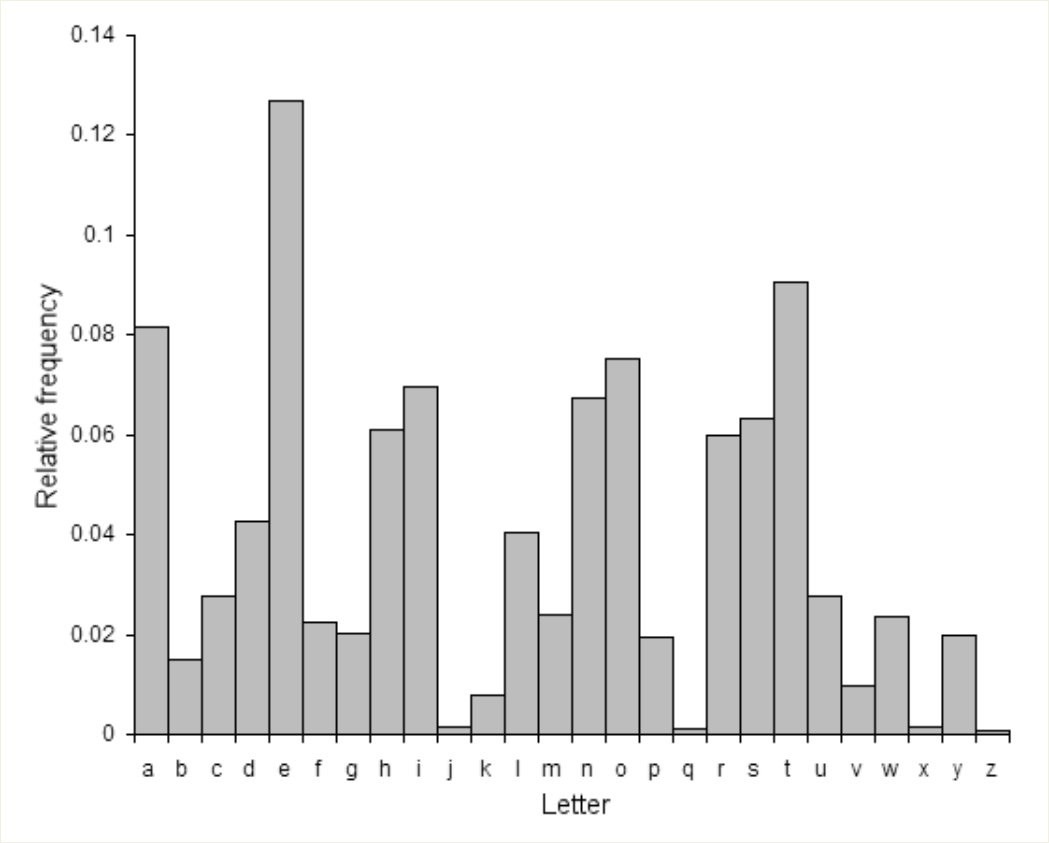
"attackatdawn"  "dwwdfndwgdzq"

Problem: Frequency of letters

Letters with a high rate of occurrence correspond to letters encrypted with high frequency also

Monoalphabetic substitution Ciphers

- The alphabet A..Z corresponds to a permutation of the alphabet
- It is still possible to break the cipher by statistical analysis of the ciphered text. For example, in English:

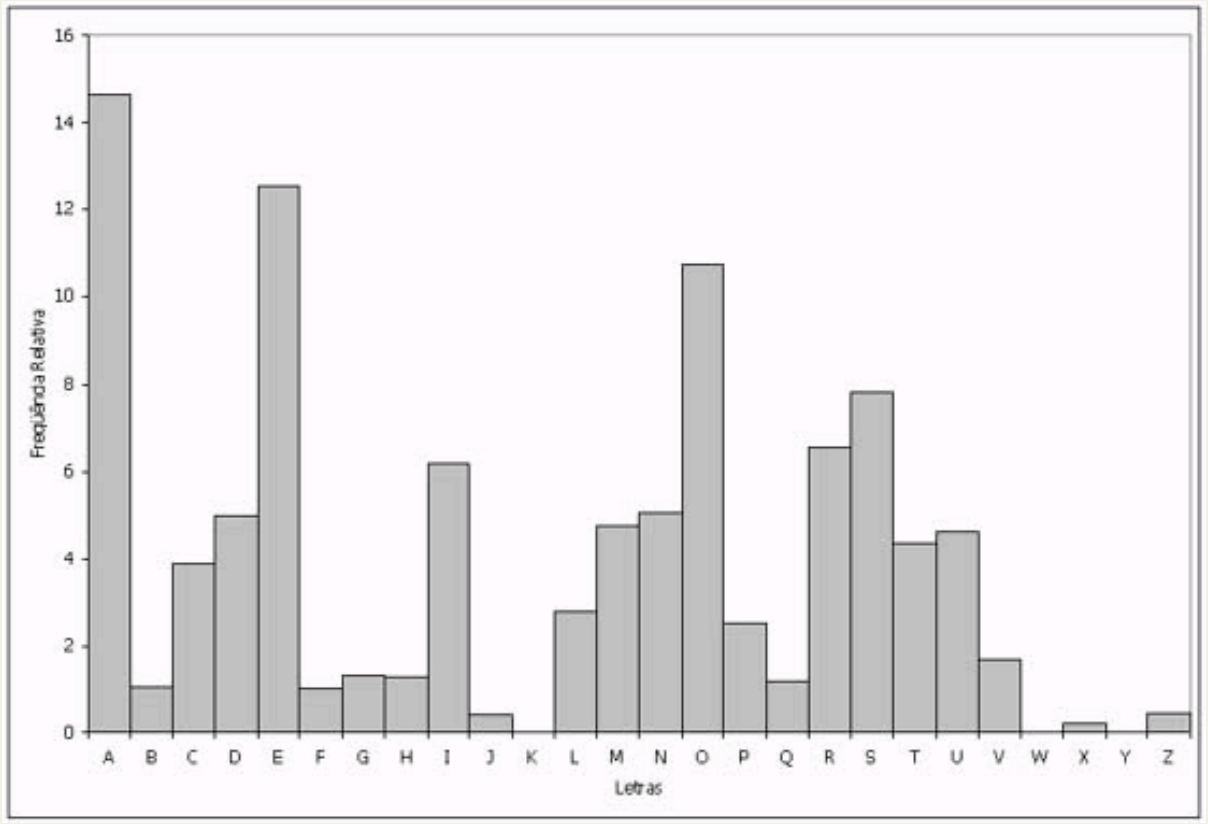


In Portuguese

A
E
O
S
R
I
D
N
T
M
U
C
P
L
V
G
F
B
H
Q

Monoalphabetic substitution Ciphers

Distribution of letters of the Portuguese alphabet:



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

Polialphabetic substitution Ciphers

- The shift applied to the letter in the original text depends on the position of the letter in the text.
- Example: If we have a text c1, c2, c3, ... ci
  - If i is divisible by 4 shift 7 letters
  - If i gives a remainder of 1 shift 5
  - If I gives a remainder of 2 shift 13
  - If i gives a remainder of 4 shift 2
- The shifts can be encoded in a word

Vigenère cipher

- Encode letters as numbers (A=1, B=2, etc.)
- Key is a “keyword”
- Cipher method
  - Add a "key word" to original text (letter by letter)
- Decipher method
  - Subtract the "key word" to the ciphered text
- Example

```
   wearediscoveredsaveyourself
 + deceptivedeceptivedeceptive
 ----------------------------

   ZICVTWQNGRZGVTWAVZHCQYGLMGJ
```

Break the Vigenère cipher

- Babbage (1854), Kasiski (1863)
- Determine length of key K, provided they knew the language used in clear text (original), and a reasonable amount of ciphered text.
- Based on search repeated encrypted strings; correspond to common groupings of two or three letters (example: the, …)
  - The same encoded sequence almost certainly corresponds to a coincidence of the same values in original text with the same key portions
  - This determines the key length and key parts ...

“One Time Pads”

- Similar to the Vigenère cipher but where the value to add is a random sequence of infinite length
- The sequence to add is part of a book in which each page is a sequence of random numbers
- The sender and receiver combine the page to be use and after this is not used again. The sender and receiver have two notebooks (pads) that contain the same sequence of numbers and never reuse and never returns to the beginning (one-time pad)
- Is a safe method against statistical analysis
- The problem is the security and maintenance of books

“One-Time Pads”

- Can be accomplished using a random number generator with the same seed (SEED)
- The “exlusive-or” (XOR) is useful in this context
- P(i): element i of the original text
- C(i): element i of the encrypted text
- O(i): value used to cipher the element i

Cipher function C(i) = P(i) XOR O(i)
Decipher function P(i) = C(i) XOR O(i)

(A xor B) xor B = A

XOR

Ent 1	Ent 2	Saída
0	0	0
0	1	1
1	0	1
1	1	0

Property: (c XOR k) XOR k = c

Using the same sequence of numbers to cipher and decipher.

“One-Time Pad”

Cipher function

```
srand (KEY)
while (text to cipher)
    c[i] = p[i] xor rand()
```

Decipher function

```
srand (KEY)
while (text to decipher)
    p[i] = c[i] xor rand()
```

“KEY” is a combined value between sender and receiver (secret key)

24

Transposition cipher

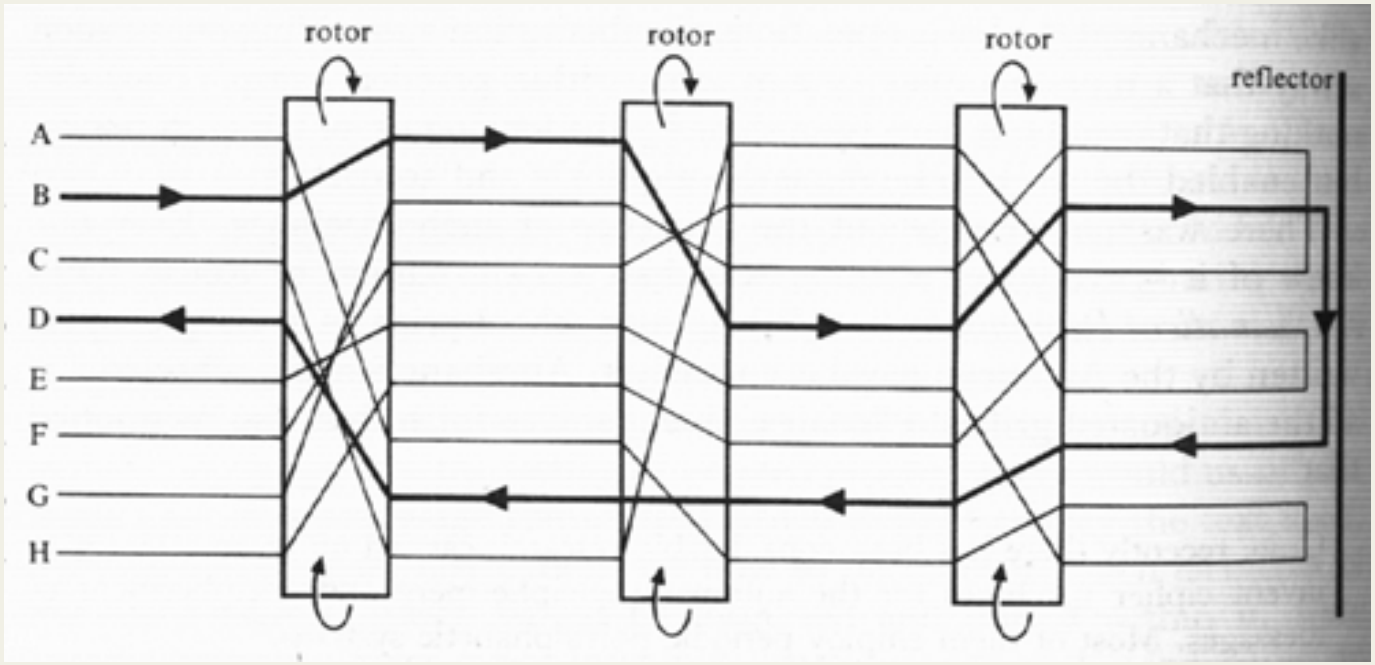
- The ciphered message is a permutation of letters in the original message. The message is divided into blocks of size N. Within each block letters are rearranged
- Suppose a block size of 5 each and the permutation is specified by the sequence (4 3 1 5 2): the 1st letter passes to 4th, the 2nd letter passes to 3rd, ...
- Example:

UM TESTE A CIFRA DE BLOCO	(word)
UMTES TEACI FRADE BLOCO	(blocks)
TSMUE AIETC AERFD OOLBC	(transposition)

Transposition cipher

- By itself, the transposition ciphers are easy to break
- Correspond to solve anagrams
- They can be used in conjunction with substitution ciphers
- Many modern encryption systems are based in the conjunction o substitution cipher with transposition cipher

Enigma Machine



Simon Singh, The Code Book, 1999,
Fourth Estate, London

There are several simulators of Enigma on the Internet
http://homepages.tesco.net/~andycarlson/enigma/enigma_j.html

Encryption and Information theory

Perfect secrecy – Shannon

An encryption system ensures ***perfect secrecy*** when: "given an encrypted message c it corresponds to the probability of a given message m has been generated with a key k is equal to the probability of m"

Observing the encrypted bytes we do not get any information about the bytes in the original file.

For this to happen it is necessary that:

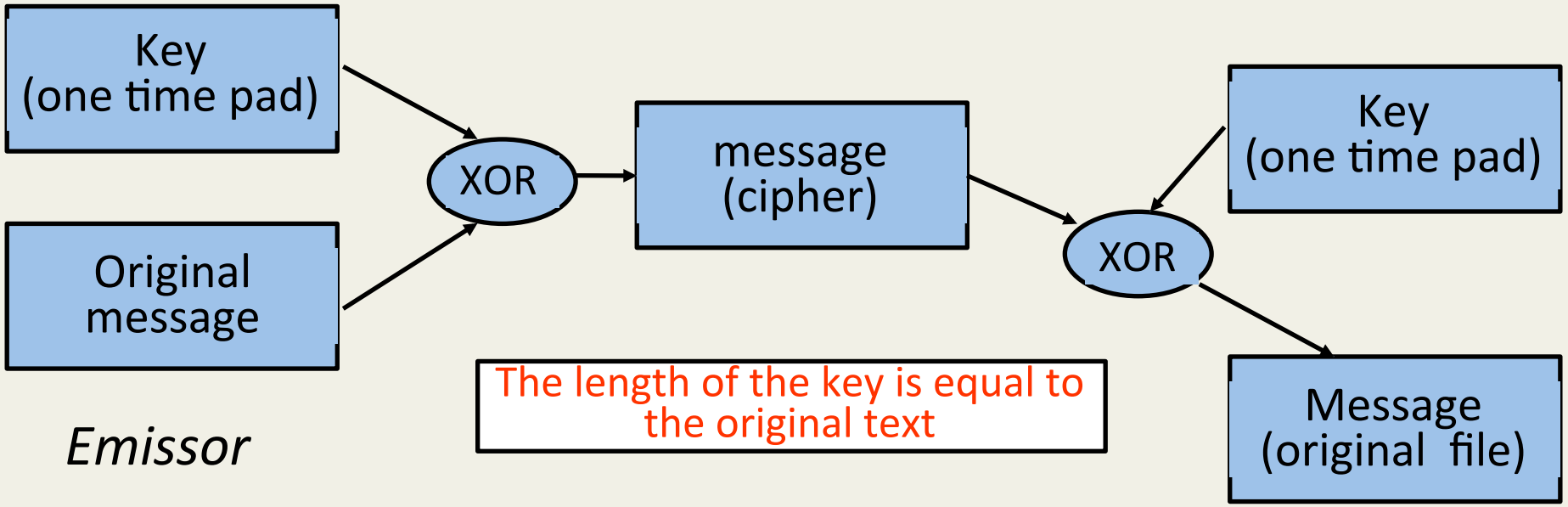
- the number of possible keys must be greater than the number of possible texts in course possible
- he choice of random keys is to make all keys equiprobable

The choice of random keys is to make all keys equiprobable

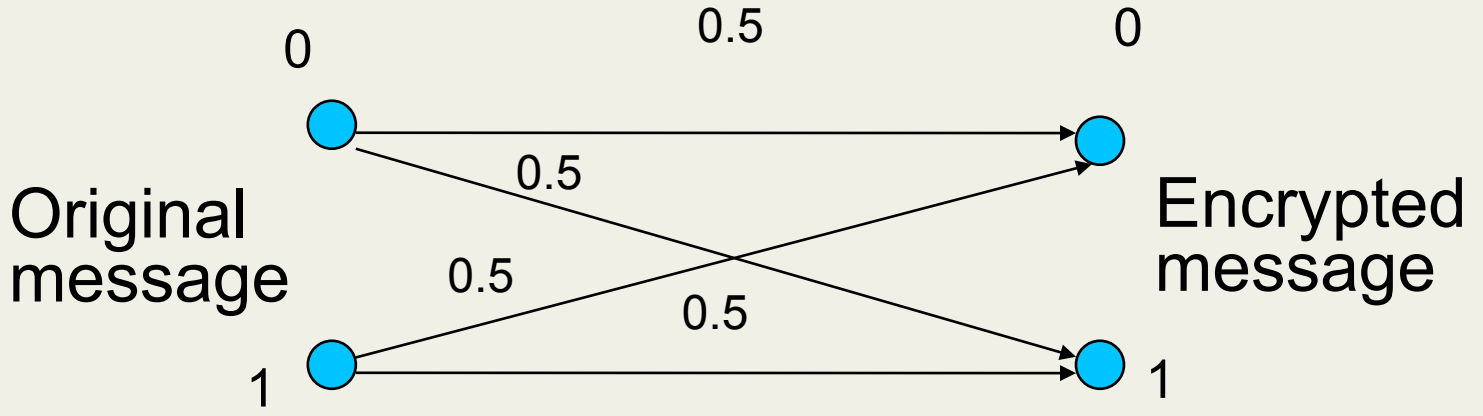
"One time pad" is a perfect cipher

As the number of possible keys is usually limited, the ciphers usually are not perfect.

An example of a perfect cipher is the "one time pad" (or Vernam cipher).



The perfect segret and the (Binary Symmetric Channel)



Error probability is associated with the key:

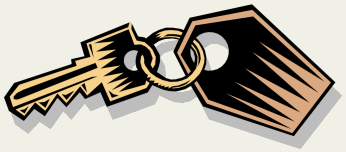
- if 0 stays equal;
- If 1 exchange.

Error probability is 0.5

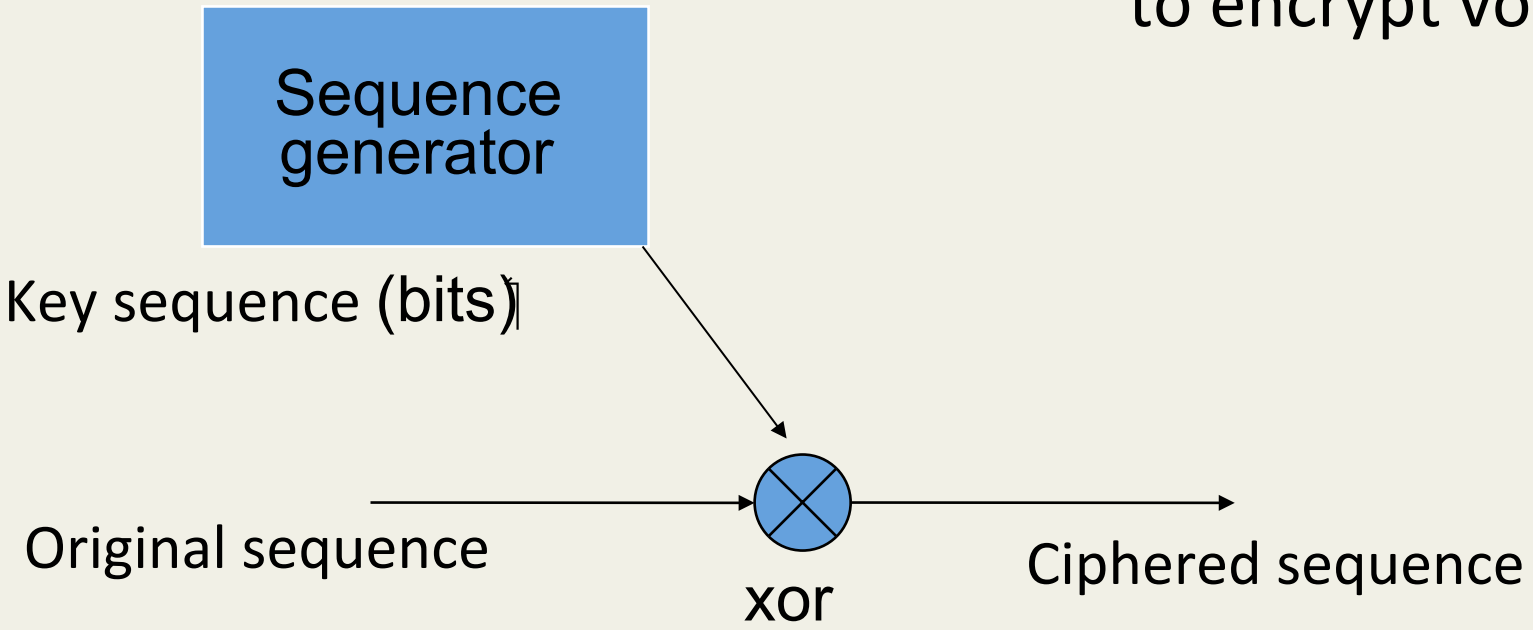
Channel capacity is 0, ie conditional uncertainty of original text (input channel) given the ciphered text (output) is the same as the uncertainty of the ciphered text

The original text and the ciphered text are independent random variables, so there is perfect secret

Stream ciphers ("contínuas")



Used for example to encrypt voice



Note: see the XOR example ("one time pad")

Symmetric stream ciphers
Generators

- Deterministic state machines, controlled by a key of finite dimension. The key determines:
  - Initial state
  - Parameterize the function that defines the next state
- It produce a cyclic sequence of bits
- The principle of confusion is used, since there is a complex relationship between the key bits in clear (original) and ciphered bits
- The principle of diffusion is not used
- Thus, typically it uses a different key for each interaction

Why is not this method always used?

The key management is not practical:

- For each text a different key has to be used;
- The keys must have length greater than or equal to the message;
- Their use in communication involves a pre-distribution of keys with a large length (size);
- It makes no sense to use this method to encrypt stored data.

Shannon criteria to assess the quality of a imperfect cipher (1)

- Amount of secrecy offered
  - Minimum security time of the cipher text, given the effort and money invested in their cryptanalysis
- Key size
  - Inherent complexity of the transmission and safeguarding of keys
- Simplicity of realization and exploitation
  - Ease of use of the cipher in production environments (criterion not relevant nowdays - software or hardware encrypt and decrypt)

Shannon criteria to assess the quality of a imperfect cipher (2)

- Error propagation
  - Undesirable because it requires retransmissions
  - Desirable because it facilitates the verification of the integrity of ciphered text
- Dimension ciphered text
  - Dimension less than or equal to the original message

Diffusion and confusion

Approximations defined by Shannon to make ciphers with a good amount of secrecy

Confusion

- The relationship between the original text, the key and ciphered text should be as complex as possible; discover parts of clear text should be difficult even when knowing part of it;
- It must be very difficult to deduce the used key used from the ciphered text.

Diffusion

- Each piece of the encrypted message should depend on a large piece of the original message;
- Each bit of the ciphered text should influence many bits of encrypted message;
- Any small change in the original message clear leads to major changes in the ciphertext

Stream ciphers generators

The keys of the stream ciphers must be as close as possible to the “one time pad”

- Longest period possible and if possible exceeding the size of the message to encrypt

- Truly random sequence:
  - equal probability
  - unpredictability

Good encryption practices

It must be assumed that the crypto-analyst:

- Knows the encryption algorithm used and its possible weaknesses, so the security is based only on the lack of key;
- Has access to all produced encrypted messages using a given algorithm and a given key;
- Knowns portions of the original message, and can use them to perform known content-based attacks.

Information Coding
2013/2014

Steganography

Steganography

- From Greek *steganós* (hidden) + graphy
- Hide information in a document
  - Text
    - Apparently neutral's protest is throughly discounted and ignored. Isman hard hit: Blockade issue affects pretext for embargo on by-products, ejecting suets and vegetable oils
    - Pershing sails form NY June 1
  - Image bitmapped: exchange the least significant bit of the bytes R, G e B; the human eye does not notice the difference
  - Idem for WAV files

Stools and OpenStego

- Stools
  - Andy Brown 1996
  - Multiple files in the same image, files can be compressed
  - Paraphrase is used to generate a key, the encryption algorithm used is symmetric (IDEA, DES, ...)
  - The MD5 algorithm is applied to the paraphrase to generate one sequence of (pseudo) random numbers that indicate the bits that contain the information
  - Demonstration S-Tools:

    http://www.cs.vu.nl/~ast/books/mos2/zebras.html
- OpenStego http://openstego.sourceforge.net/

Examples

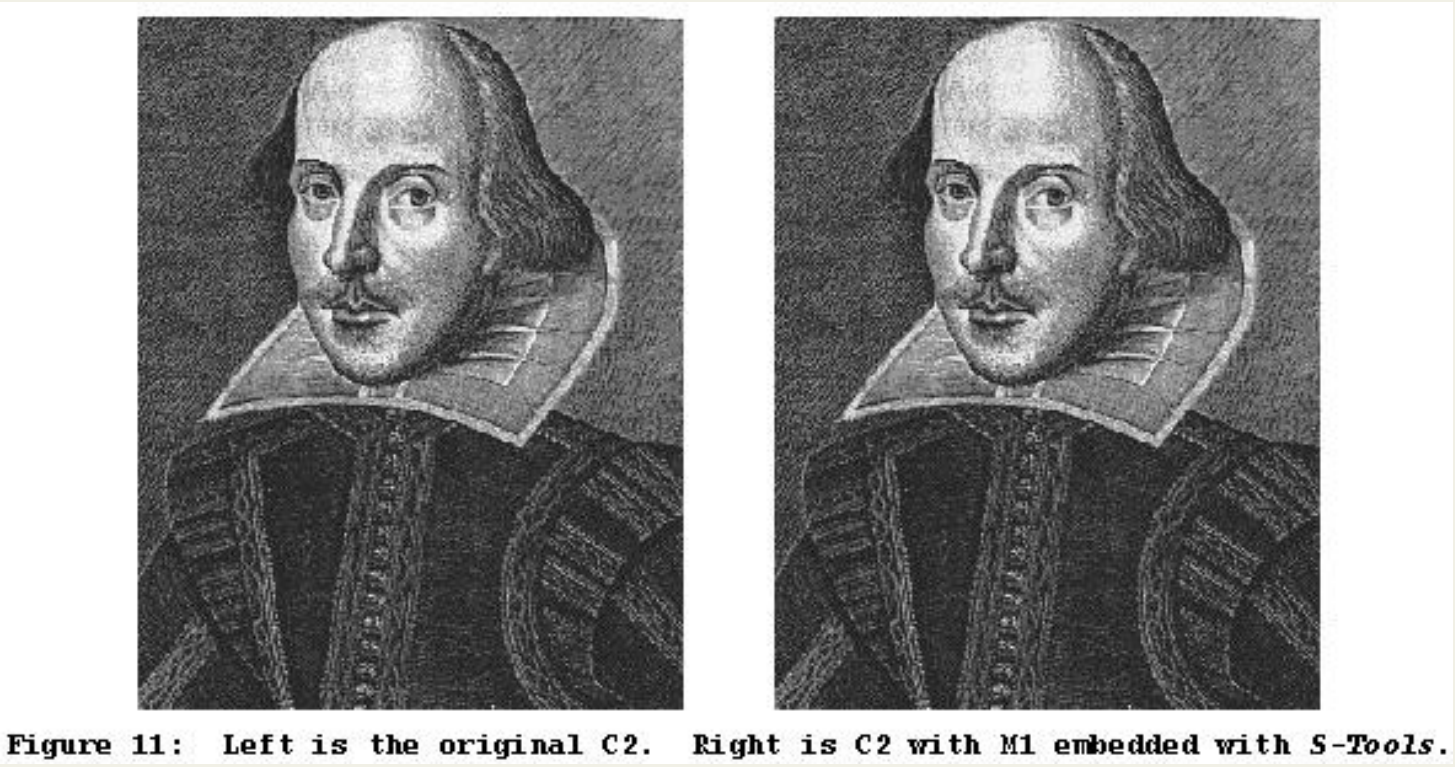


Figure 11: Left is the original C2. Right is C2 with M1 embedded with *S-Tools*.

The added text is 518 bytes

Examples

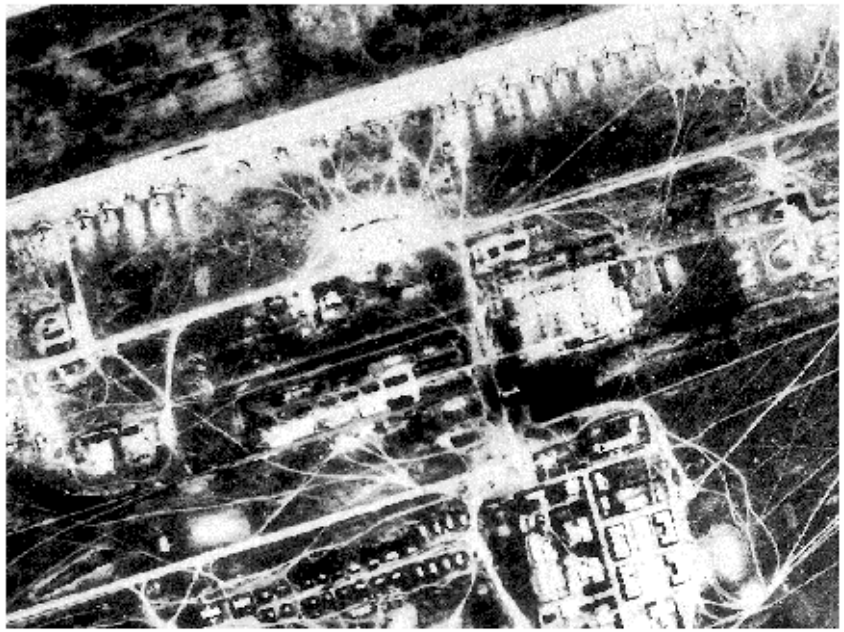
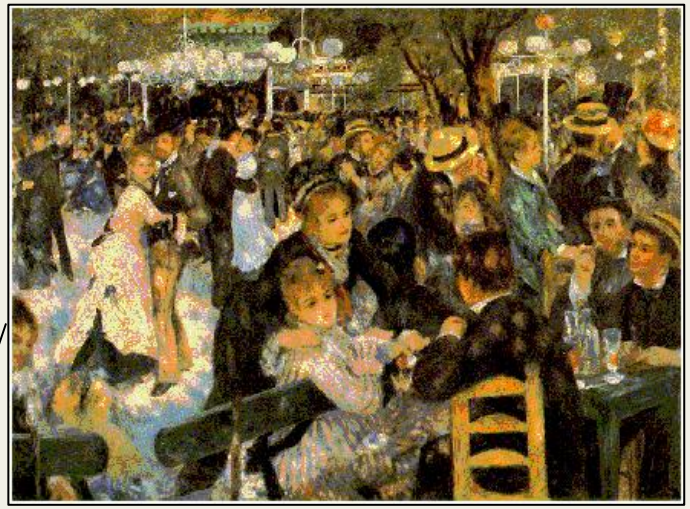
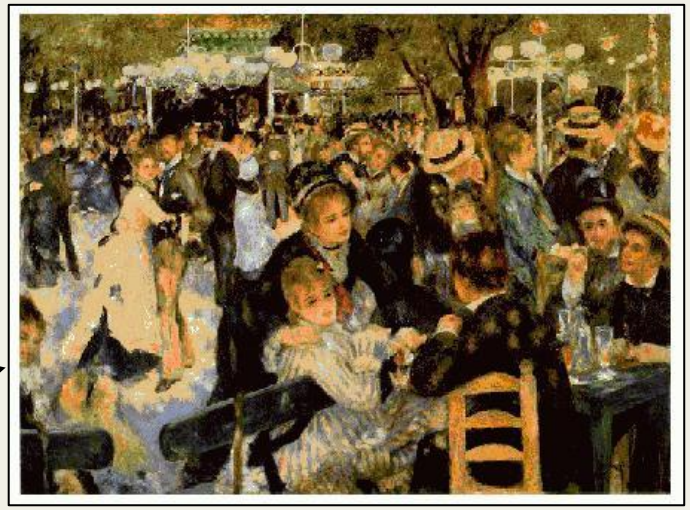


Figure 4: Long-Range Aviation Airfield[5]

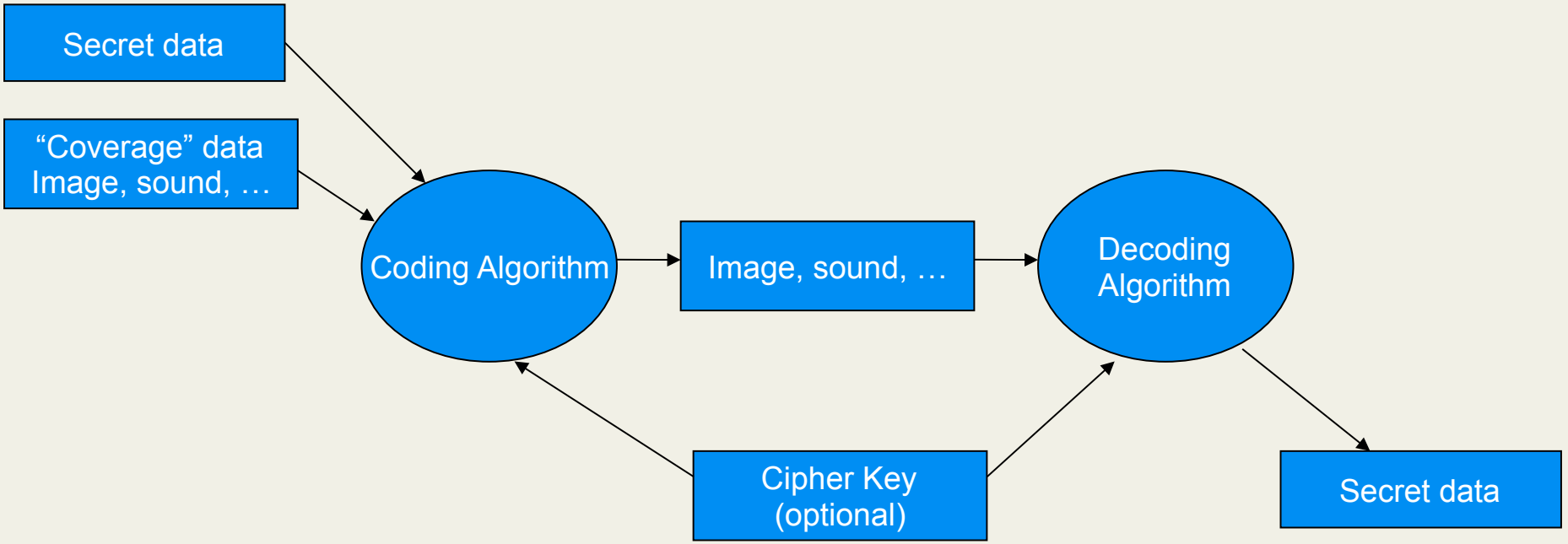


A.Renoir
Moulin
De la
Gallete

Stools



Steps to hide and recover information



Steganography in images

- Regular spread
- Placement in areas with noise in order to “disguise”
- Using a sequence of random numbers to define which bits to change
- Change the least significant bit
- Preferable in lossless formats (BMP)
  - 24 bits preferable to 8 bits
  - Problematic when there is a LookupTable (LUT): GIF, TIFF
  - On average half the bits must be changed

00100111 11101000 11001000

00100110 11001000 11101000

11001000 00100111 11101001

One example (1)

Steganography works by changing the value of the pixels in the original image, where some bits of the image pixels are selected to "hide" a message. The changes in the values of the pixels should not cause visible changes in the initial image. Who gets altered image shall recover the message by extracting the respective bits of each pixel.

One example (2)

1. In the first two pixels of the original image (eg in ppm format) is saved the size of the file that contains the message you want to "hide". Each pixel is represented by 3 bytes, each corresponding to one of the color components (r, g, b).

   (Assuming that the message file size does not exceed the capacity of 2 bytes).

One example (3)

2. The bytes of the message file is "hidden" in the pixels of the image by working the following pattern for each byte:
   - Pixel: r7 r6 r5 r4 r3 r2 r1 r0 , g7 g6 g5 g4 g3 g2 g1 g0 , b7 b6 b5 b4 b3 b2 b1 b0
   - Message byte: m7 m6 m5 m4 m3 m2 m1 m0
   - Modified pixel: r7 r6 r5 r4 r3 m7 m6 m5 , g7 g6 g5 g4 g3 m4 m3 m2 , b7 b6 b5 b4 b3 b2 m1 m0
3. That is, each pixel "hide" one message byte.

One example (4)

- Example: considering that we have pixel <225, 100, 100> and we want to hide ‘a’:
  - Pixel: 1 1 1 0 0 0 0 1 , 0 1 1 0 1 0 0 0 , 0 1 1 0 1 0 0 0
  - ‘a’: 0 1 1 0 0 0 0 1
  - Modified pixel: 1 1 1 0 0 0 1 1 , 0 1 1 0 1 0 0 0 , 0 1 1 0 1 0 0 1
- The modified pixel has the value (227, 104,105), as you can see not much different from the original pixel value.

One example (5)

- The selection of what pixels to modify can be uniform across the entire image, taking into account the size of the original file and message file. Taking into account, of course, that the embedded messages can no be larger than the number of pixels of the image.