Resum IS

Informació i Seguretat (Universitat Autònoma de Barcelona)

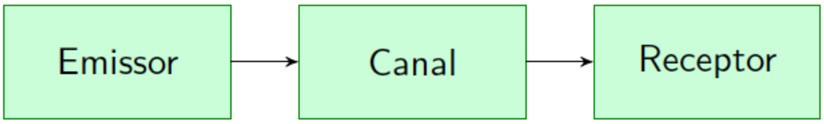
INFORMACIÓ i SEGURETAT

Índex de continguts

Tema 1

Motivació. Planteig dels problemes de la comunicació

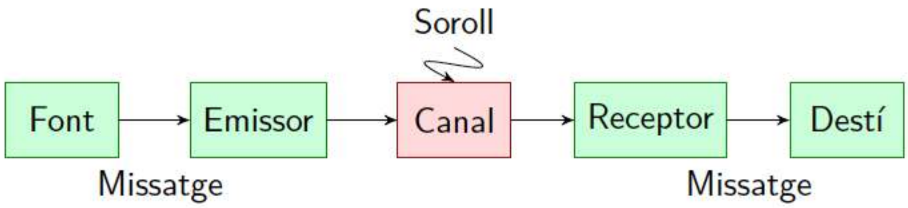
Model d'un sistema de comunicació



Problema de la comunicació

Eficient: Minimitzant recursos (espai i temps).
Exacta: Sense pèrdua d'informació.
Segura: Sense manipulació de la informació.

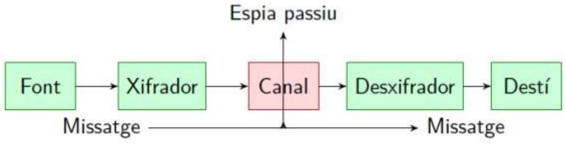
Arquitectura d'un sistema de comunicació



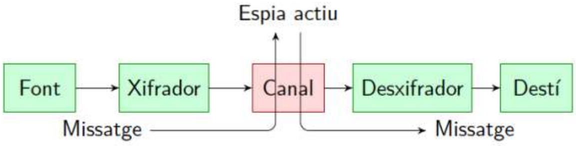
- **Font**: persona o màquina que produeix informació.
- **Emisor/Codificador**: adapta la informació a les característiques del canal (seq. dígits, polsos elèctrics,...).
- **Canal**: mitjà pel qual es transmet el missatge (espai, fil, disc,...).
- **Soroll**: error (pertorbació) aleatori característic del canal.
- **Receptor/Descodificador**: recuperació de la informació, sovint amb errors.
- **Destí**: persona o màquina que rep la informació.

Espies: privacitat i autenticitat

Privacitat: preservar la confidencialitat de les dades (espia passiu).



Autenticitat: impedir la modificació no autoritzada de les dades (espia actiu).



Tema 2

Conceptes bàsics de la teoria de la informació

La quantitat d'informació que obtenim després de realitzar l'experiment és igual a la quantitat d'incertesa abans de realitzar-ho.

$$Quantitat\ d'informació\ =\ Quantitat\ d'incertesa.$$

<u>Mesura de Hartley</u> : Incertesa sobre n resultats possibles i *equiprobables.*

$$I(n) = log(n)$$

<u>Problema</u>: no es tenen en compte les probabilitats de cada resultat

<u>Mesura de Shannon</u>: Incertesa/informació d'un esdeveniment A amb probabilitat $p(A)$

$$I(A) = \log\frac{1}{p(A)} = -\log p(A)$$

<u>Entropia</u>: És la magnitud que mesura la informació continguda en un flux de dades, o d'una font.

$$H(X) = \sum\left(p(a_i)\cdot I(A_i)\right) = \sum\left(p(a_i)\cdot\log\frac{1}{p(a_i)}\right)$$

↳ Unitats de l'entropia
bits/resultat o bé *bits/missatge*

<u>NOTA</u>: $\log_2$ → Perquè volem els resultats en bits

<u>Entropia d'una variable aleatòria</u>
Quantitat d'informació mitjana que té una font d'informació.

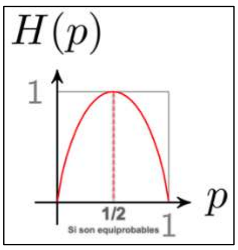
$$H(X) = -\sum(p_i\cdot\log p_i)$$

Propietats:

$$\begin{cases} H(X) \geq 0 \\ H(X) = 0\ si\ hi\ ha\ una\ p_i = 1 \end{cases}$$

$$\begin{cases} H(X) \leq \log n \\ H(X) = \log n\ \ si\ p_i = {}^{1}\!/_{n}\ \forall_i \end{cases}$$

<u>Entropia binària</u>
Entropia d'una font que emet zeros i uns.
Si tenim 2 esdeveniments:



$$S = \{a_1, a_2\}\begin{pmatrix} p(a_1) = p. \\ p(a_2) = 1-p \end{pmatrix}$$

$$H(X) = p\cdot\log\frac{1}{p} + (1-p)\cdot\log\frac{1}{(1-p)}$$

↳ $Com\ a\ màxim\ serà\ 1\ bit.$

Entropia conjunta

Tenim que X, Y son variables aleatòries discretes que representen entrada i sortida d'un canal.

$$H(X,Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_i) \cdot \log p(x_i, y_i)$$

Entropia condicionada

Donat un valor $Y = y$ obtenim X

$$H(X|Y) = -\sum_{i=1}^{n}\sum_{j=1}^{m} p(x_i, y_i) \cdot \log p(x_i|\, y_i)$$

Informació mútua

Mesura la dependència mútua de dues variables aleatòries, és a dir, la reducció d'incertesa (entropia) d'una variable aleatòria X, donat el coneixement del valor d'una altre variable aleatòria Y.

$$I(X,Y) = H(X) - H(X|Y) = H(Y) - H(Y|X)$$

Capacitat del canal

Quantitat màxima d'informació que pot passar per símbol d'entrada. Unitat de mesura bits/símbol.

$$C = \max\big(I(X,Y)\big) = \max\big(H(X) - H(Y|X)\big)$$

Tema 3
Codificació de la font

__Conjunt de missatges__:
$S = \{a_1, \dots, a_n\}$

__Alfabet D-ari del codi__
$A = \{\sigma_1, \dots, \sigma_n\}$
__Ex. Alfabet Binari/2-ari__
$A = \{0,1\}$

__Codi__
$C = \{c_1, \dots, c_n\}$

Codis de Longitud Fixa
Alguns exemples en poden ser l'ASCII o UNICODE

$$D^L \geq n \qquad L \geq \frac{\log n}{\log D}$$

$$\begin{bmatrix} L \rightarrow Longitud \\ D \rightarrow Símbols \\ n \rightarrow paraules - codi \end{bmatrix}$$

Si codifiquem seqüencia de símbols (m-tuples de missatges):

$$D^L \geq n^m \qquad L \geq \frac{m \cdot \log n}{\log D}$$

Codis de Longitud Variable
Alguns exemples en poden ser el Morse o Zip. Només treballarem amb codis de descodificació única.

$$Longitud\ mitjana = \bar{L} = \sum_{i=1}^{n} p_i\, L_i$$

Codis instantanis
Un codi és instantani si cap paraula-codi és prefix d'una altra. Tot codi instantani és de descodificació única, però no a l'inrevés.
Per comprovar si existeix un codi instantani utilitzem el teorema de *Desigualtat de KRAFT*.

$$\sum_{i=1}^{n} D^{-L_i} \leq 1$$

Primer Teorema de Shannon
Estableix el límit teòric per a la compressió d'una font de dades (origen).
Per a obtenir aquest mínim valor de la Longitud mitjana utilitzem:

$$\bar{L} \geq \frac{H(S)}{\log D}$$

Eficiència i redundància d'un codi
L'eficiència d'un codi D-ari amb longitud mitjana $\bar{L}$ per a un conjunt de missatges S és:

$$\eta = \frac{H(S)}{\bar{L} \cdot \log D}\ (\leq 1)$$

La redundància és el valor complementari de l'eficiència.

$$1 - \eta\ (\geq 0)$$

Codi òptim

Un codi és òptim si per el mateix conjunt de missatge només existeixen altres codis amb longitud mitjana $\bar{L}$ igual o superior.

- Si $\eta = 1$ aleshores és òptim.
- Si comparem codis D-aris pel mateix conjunt de missatges S
  - Comparem les longituds mitjanes $\bar{L}$
- Si comparem codis D-aris amb diferent alfabet.
  - Comparem les eficiències

Construcció de codis òptims (Mètode Huffman)

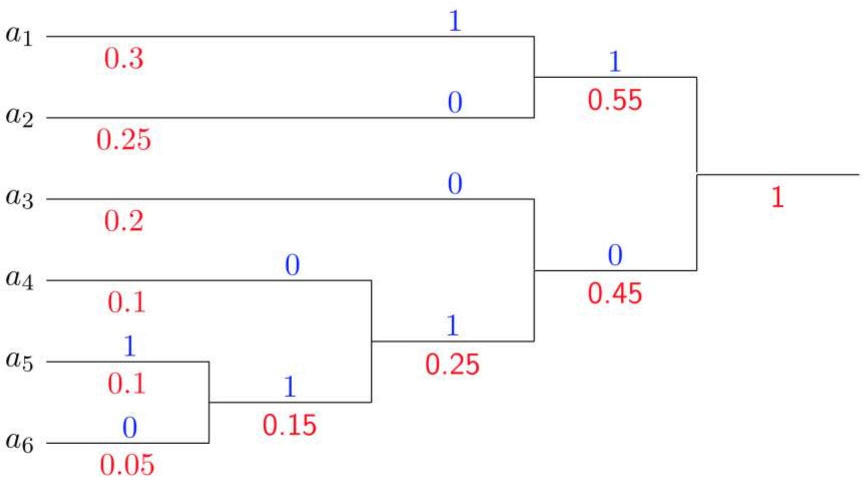
És un mètode vàlid per a qualsevol alfabet.

Algorisme:

1. Ordenem els missatges de S per ordre decreixent de possibilitats.
2. Assignem 1 i 0 com a últims símbols de les paraules-codi per als dos missatges menys probables ($a_{n-1}\ i\ a_n$).
3. Reagrupem els dos últims missatges com un nou missatge amb probabilitat $p_{n-1} + p_n$
4. Si queda més d'un missatge tornem a 1.

Exemple Huffman

L'arbre resultant representa el codi de Huffman òptim.



$$C = \{11, 10, 00, 010, 0111, 0110\}$$

Tema 4

Compressió de la font

Taxa de compressió

Suposem un fitxer M comprimit a un fitxer C, la seva taxa de compressió serà la mida del fitxer comprimit dividida entre la mida del fitxer no comprimit, on aquesta taxa estarà mesurada en bits per bit (bpb).

$$R = \frac{|C|}{|M|}$$

Percentatge de compressió

En aquest cas el tant per cent % de compressió del fitxer vindrà donat per:

$$(1 - R) \cdot 100\%$$

Bitrate

També anomenada *taxa de bits*, és la freqüència amb que les dades es transmeten en el medi. Es mesura en bits per símbol.

$$BR = \frac{|C|\ bits}{|M|\ símbols}$$

Mètodes de compressió

- **Per Taxa de Compressió**
  - *Sense pèrdua (lossless)*: És totalment reversible $I = \hat{I}$
  - *Amb pèrdua (lossy)*: No és totalment reversible $I \neq \hat{I}$
- **Per Model**
  - *Estàtics (no adaptatius)*: el model d'estimació és fix durant tot el procés de compressió.
  - *Dinàmics (adaptatius)*: el model canvia durant al procés de compressió

Tècniques de compressió

- **Basats en repetició** → RLE. Elimina redundància quan la font repeteix molts missatges
- **Mètodes estadístics** → Huffman, aritmètic. Elimina la redundància estimant les probabilitats dels missatges.
- **Basats en diccionari** → LZ, LZW. Elimina la redundància creant un diccionari de missatges més freqüents.
- **Basats en transformades**→ DCT, Wavelets, JPG. Transforma missatges per descorrelacionar les dependències estadístiques.

$$Cal\ mirar\ els\ procediments\ dels\ exercicis\ per\ a\ saber\ fer:$$
$$\boldsymbol{RLE}, \quad \boldsymbol{Codificació\ Aritmètica}, \quad \boldsymbol{Lempel - Ziv (LZ77\ i\ LZ78)}$$

Compressió d'Imatges

- Mètodes més usuals
  - Sense Compressió: BMP, RAW, PGM, PPM
  - Sense Pèrdua: PNG, GIF, TIFF
  - JPEG
  - JPEG2000
  - MPEG

Tema 5

Codificació del canal

Tipus de Canals

- Canal Sense Pèrdua
  - $H(B|A) = 0 \rightarrow I(A,B) = H(A)$
  - $n \leq m$
  - Cada columna de Π té un únic element $\neq 0$
  - $C = \max_{\{p_i\}_{i=1}^{n}} H(A) = \log n$

Matriu del canal:

$$\Pi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix},$$

$C = \log 2 = 1$ bits/símbol entrada

- Canal Determinista
  - $H(B|A) = 0 \rightarrow I(A,B) = H(B)$
  - $m \leq n$
  - Cada fila de Π hi ha un únic element $\neq 0$
  - $C = \max_{\{p_i\}_{i=1}^{n}} H(B) = \log m$

Matriu del canal:

$$\Pi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{pmatrix},$$

$C = \log 2 = 1$ bits/símbol entrada

- Canal Sense Soroll
  - És sense soroll si és *Sense Pèrdua* i *Determinista*
  - $I(A,B) = H(A) = H(B)$
  - A cada columna de Π hi ha un 1 i la resta 0
  - $n = m$
  - $C = \log n = \log m$

Matriu del canal:

$$\Pi = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

$C = \log 2 = 1$ bits/símbol entrada

- Canal Totalment Simètric
  - Les files i les columnes de Π són iguals excepte canvi d'ordre.
  - $C = \log m - H$ on $H = H(B|A)$

Matriu del canal:

$$\Pi = \begin{pmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} \end{pmatrix},$$

$C = \log 4 - H(\frac{1}{3}, \frac{1}{3}, \frac{1}{6}, \frac{1}{6}) = 0.081$ bits/símbol entrada

- Canal amb Entrada i Sortida Independents
  - $H(A|B) = H(A)$ i $H(B|A) = H(B)$
    ↳ $I(A,B) = 0$
  - No serveix per transmetre informació
  - Les files de Π són idèntiques
  - $C = 0$

Matriu del canal:

$$\Pi = \begin{pmatrix} \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

$C = 0$ bits/entrada

Probabilitat mitjana d'error

$$\overline{p_e} = \sum_{j=1}^{m} p(B_j) p_e(B_j) = 1 - \sum_{j=1}^{m} p(f(B_j), B_j)$$

Regla MPE (Mínima Probabilitat d'Error)

Aquesta regla minimitza la probabilitat mitjana d'error.

Regla a MV (Màxima Versemblança)

En un canal amb distribució inicial de probabilitat equiprobable, les regles MPE i MV coincideixen.

Exercicis que cal mirar:

MPE en un BSC   BSC – Codificació de la font   BSC – Regla MPE o MV

BSC – Probabilitat mitjana d'error   BSC – Taxa de transmissió de la informació

Tema 6

Teoria de la Codificació per a la correcció d'errors

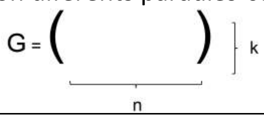
Símbol	*Descripció*
$\mathbb{F}$	Alfabet del canal, en codis binaris $\mathbb{F} = \{0,1\}$
$\boldsymbol{v,u}$	Paraula, és una seqüència de dígits binaris $v = (v_1 v_2 \cdots v_n) \in \mathbb{F}^n$
$\boldsymbol{n}$	Longitud de la paraula
$\boldsymbol{C}$	Codi Binari, és un subconjunt de paraules de longitud n, és a dir $C \subset \mathbb{F}^n$
$\boldsymbol{M}$ **o** $\boldsymbol{\lvert C \rvert}$	Número de paraules
$\boldsymbol{R_T}$	Taxa de transmissió de la informació $R_T = \frac{\log_2(\lvert C \rvert)}{n}$ bits/símbol
$\boldsymbol{r}$	Redundància $r = 1 - R_T$
$\boldsymbol{p}$ // $\boldsymbol{\overline{p}}$	Probabilitat d'error // Probabilitat mitjana d'error
$\boldsymbol{d_H}$	Distància Hamming. $d_H(u,v) = \lvert\{i \mid u_i \neq v_i, 1 \leq i \leq n\}\rvert$
$\boldsymbol{d}$	Distància mínima. $d = min\{d_H(u,v) \mid u \neq v,\ u,\ v \in C\}$

CODI BINARI $C[n,M,d]$		
$C[3,2,3]$ $C[6,4,3]$	$C = \{000, 111\}$ $(v = 000\ u = 111)$ $C = \{000000, 010101, 101010, 111111\}$	$n = 3,\ M = 2,\ d = 3$ $n = 6,\ M = 4,\ d = 3$

Símbol	*Descripció*
$\boldsymbol{t}$	Capacitat Correctora $t = \left\lfloor \frac{d-1}{2} \right\rfloor$ (←No agafem decimals) $t = \left\lfloor \frac{4-1}{2} \right\rfloor = \left\lfloor \frac{3}{2} \right\rfloor = 1$
$\boldsymbol{\delta}$	Capacitat Detectora $\delta = \mathrm{d} - 1$

*Sigui **C** un codi binari de distància mínima* $\boldsymbol{d}$		
Podem detectar $\boldsymbol{r}$ errors	*si, i només si*	$\boldsymbol{r} < \boldsymbol{d}$
Podem corregir $\boldsymbol{t}$ errors		$2\boldsymbol{t} < \boldsymbol{d}$
Podem corregir $\boldsymbol{s}$ esborralls		$\boldsymbol{s} < \boldsymbol{d}$
Podem corregir $\boldsymbol{t}$ errors i $\boldsymbol{s}$ esborralls		$2\boldsymbol{t} + \boldsymbol{s} < \boldsymbol{d}$

Propietats dels codis binaris lineals

Símbol	***Descripció***
Codi Lineal $\boldsymbol{C[n,k,d]}$	Codi binari C on per tota parella $u, v \in C$ aleshores $u + v \in C$. És a dir, $C \subset \mathbb{F}^n$ és un subespai vectorial de $\mathbb{F}^n$
Per a que un codi sigui lineal: • Ha de tenir la paraula tot zeros. $\mathbf{0} = (0 \dots 0)$ • Ha de tenir 2^k paraules-codi. $M = \lvert C \rvert = 2^k, \ k \leq n$ • La suma de tota parella de paraules codi ha de ser una altra paraula codi. $(u + v) \in C$ (Suma és bit a bit)	
$\boldsymbol{k}$	<u>Dimensió</u> de C com a subespai vectorial de $\mathbb{F}^n$ $k = \log(\lvert C \rvert)$
$\boldsymbol{d}$	Pes mínim, és a dir, nombre mínim d'1s.
<u>Producte escalar de dos vectors</u> $(u \cdot v)$ $\boldsymbol{u} = (\mathbf{1}, \mathbf{0}, \mathbf{1}) \ \boldsymbol{v} = (\mathbf{1}, \mathbf{1}, \mathbf{0}) \rightarrow (1x1) + (0x1) + (1x0) = \boldsymbol{u} \cdot \boldsymbol{v} = \mathbf{1}$ $\boldsymbol{u} = (\mathbf{1}, \mathbf{0}, \mathbf{1}, \mathbf{0}) \ \boldsymbol{v} = (\mathbf{0}, \mathbf{1}, \mathbf{0}, \mathbf{1}) \rightarrow (1x0) + (0x1) + (1x0) + (0x1) = \boldsymbol{u} \cdot \boldsymbol{v} = \mathbf{0}$	
$\boldsymbol{C^{\perp}}$	<u>Codi Ortogonal</u>, producte escalar de dos paraules = 0 $(u \cdot v = 0)$ Tots els vectors amb número parell d'1s és ortogonal amb ell mateix. $C^{\perp}[n, n-k, d']$
$\boldsymbol{C'}$	<u>Codi Estès</u>, $C[n+1, \ k, \ d^*]$
$\boldsymbol{G}$	<u>Matriu generadora</u> del codi C, en què les k files de la matriu formen una base de C. (Una base de C son diferents paraules-codi linealment independents) 
$\boldsymbol{H}$	<u>Matriu de control</u> $G = (Id \mid x) \ \rightarrow \ H = (x^t \mid Id)$ $G = (x \mid Id) \ \rightarrow \ H = (Id \mid x^t)$ És probable que sigui necessari aplicar transformacions lineals a la nostra G perquè aquesta quedi amb la Identitat.

CODI BINARI LINEAL $C[n,k,d]$		
$C[3,2,1]$	$C = \{000, 001, 010, 011\}$	$n = 3, \ k = 2, \ d = 1, \ R_T = \frac{2}{3}$

CODIFICAR AMB UN CODI LINEAL $C[5,3,2]$		
<u>Missatge</u>	*<u>Matriu Generadora</u>*	
$\boldsymbol{u} = (101)$	$\boldsymbol{G} = \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix}$	$\boldsymbol{u} \cdot \boldsymbol{G} = (101) \cdot \begin{pmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 1 \end{pmatrix} = (10011)$

LA SÍNDROME	
Donat $w \in \mathbb{F}^n$, el vector $\boldsymbol{H}(\boldsymbol{w}) = \boldsymbol{w} \cdot \boldsymbol{H}^{\boldsymbol{T}} \in \mathbb{F}^{\boldsymbol{n-k}}$ s'anomena síndrome de w.	
$\boldsymbol{e}$	<u>Vector d'error</u>. Si enviem $\boldsymbol{v} \in C$ i rebem $\boldsymbol{w} \in \mathbb{F}^n$, el vector d'error és $\boldsymbol{e} = \boldsymbol{w} - \boldsymbol{v}$

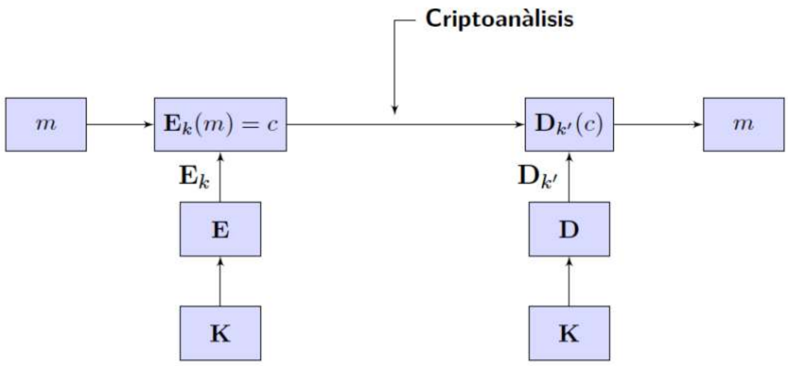
CALCULAR DISTANCIA MÍNIMA A PARTIR DE LA MATRIU DE CONTROL
Si hi ha columna tot 0's $d = 1$, sinó: Si hi ha 2 columnes iguals $d = 2$, sinó: Si la suma de 2 columnes ens dóna una altra (3 dependents) $d = 3$, sinó: Si la suma de 3 columnes dóna una altra (4 dependents) $d = 4$.

Tema 7

Criptografia bàsica i seguretat computacional

Criptografia de Clau Pública

Mètode de xifratge en el qual les claus per xifrar i desxifrar no són deduïbles fàcilment una de l'altra.

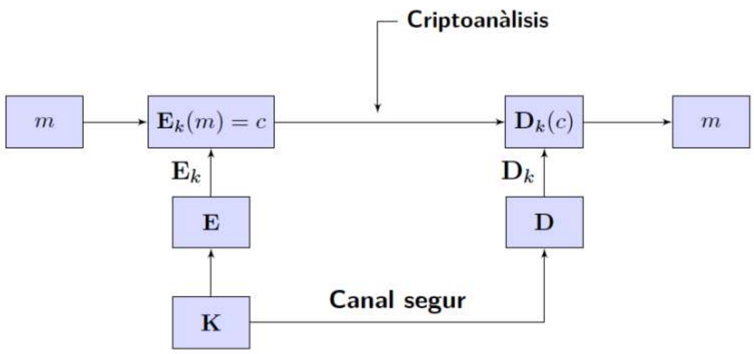


Criptografia de Clau Simètrica (Compartida o Privada)

Mètode de xifratge en el qual els claus per xifrar i desxifrar (que comparteixen emissor i receptor) són fàcilment deduïbles una de l'altra.

S'anomenen també mètodes de clau compartida o secreta.

$$E_k(m) = c, \qquad D_k(c) = m$$
$$D_k\big(E_k(m)\big) = m, \qquad E_k\big(D_k(c)\big) = c$$



Aquest tema és millor aprendre'l fent exercicis.