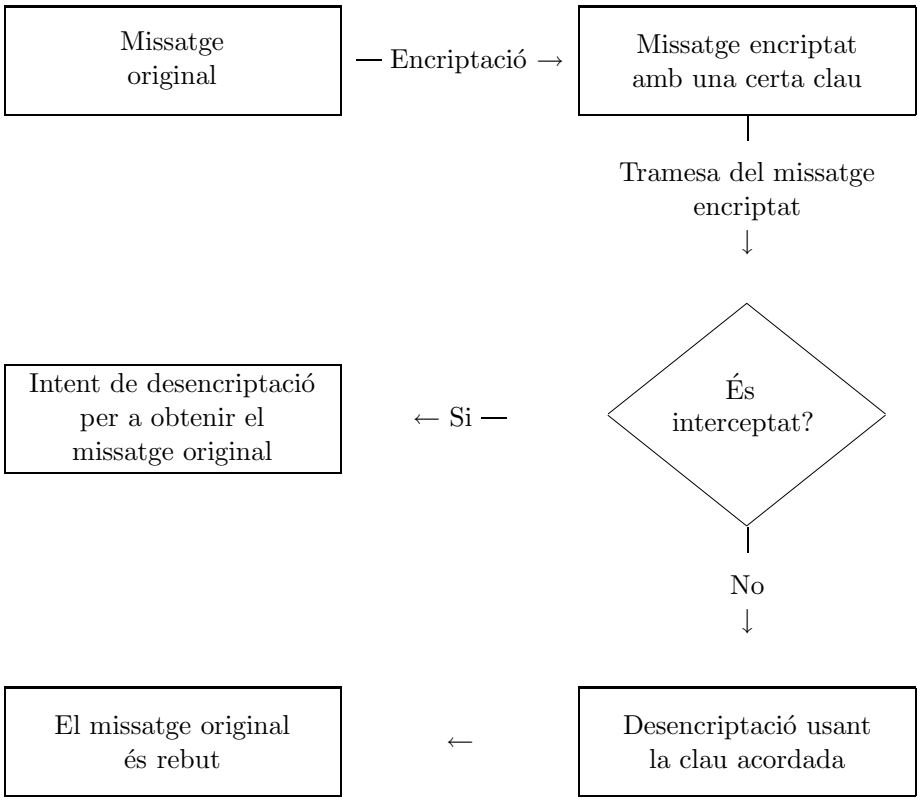
2.15 Codis secrets

La transmissió de missatges entre dues persones, de manera que no puguin ser entesos per una tercera persona, és una pràctica que ha estat usada per la humanitat des de fa més de dos mil anys. Es tenen notícies de missatges encriptats de l'època de la Grècia clàssica. Un esquema del mètode general que és el següent



Encara que durant la història hi ha hagut nombrosos xifrats (mètodes d'encriptació de missatges), hi ha dues famílies de xifrats clarament diferents: Els xifrats de substitució i els de transposició. A continuació parlarem una mica de cadascun d'aquests tipus.

Xifrats de substitució

Un dels xifrats més antics és atribuït a Juli Cèsar i ens donarà una primera idea del que són aquests tipus de xifrats. Consisteix en el següent:

Donat un missatge substituïm cadascuna de les seves lletres per la lletra situada quatre llocs més endavant en l'ordre alfabètic (quan s'acaben les lletres tornem a començar per la A). Així, per exemple, la codificació de CADA seria FDHD. En general, cada lletra es substituiria per la de sota de la següent llista.

A B C Ç D E	...	V W X Y Z
D E F G H I	...	Z A B C Ç

De seguida els desencriptadors (persones especialitzades a tractar de veure quin és el missatge original a partir del missatge codificat) van ser capaços de desencriptar missatges

encriptats per aquest mètode. Observem que desencriptant una lletra, se'n dedueixen totes les altres.

Al segle *XV*, L.B. Alberti va començar a perfeccionar el xifrat de J. Cèsar. La millora va consistir a prendre a la segona fila una permutació (reordenació) qualsevol de totes les lletres de l'alfabet. Per exemple,

$$\begin{array}{cccccccc} \text{A} & \text{B} & \text{C} & \text{Ç} & \text{D} & \text{E} & \dots & \text{X Y Z} \\ \text{H} & \text{L} & \text{A} & \text{W} & \text{T} & \text{S} & \dots & \text{I U B} \end{array}$$

D'aquesta manera s'havien de desencriptar totes les lletres una per una. Els desencriptadors ho tenien una mica més difícil, però van ser capaços de nou de trobar un sistema per a obtenir el missatge original. Aquest es basaba en estudis estadístics sobre quines eren les lletres que sortien més sovint, els grups de dues lletres que sortien més sovint, els de tres lletres, etc. A partir d'aquí, i si el missatge era una mica llarg, podien "traduir-lo". Per exemple en català tenim:

- lletres més freqüents per ordre: E, A, S, R, I, N, T, O,..., K, W.
- grups de dues lletres més freqüents: ES, AR, EN, RE, RA, SS, RI, EM, ...

i en castellà:

- lletres més freqüents per ordre: E, A, O, S, R, I, N,..., Z, K, W, X.
- grups de dues lletres més freqüents: ES, EN, EL, DE, LA,....

Per tant, en un text que prové del castellà, per exemple, és molt possible que la lletra que surti més sovint sigui la que correspon a la E. A partir d'aquesta, tenim lletres candidates a ser la S, la N o la L i així successivament.

Al segle *XVI* va haver-hi una millora important deguda a B. Vigenère. Aquesta consisteix en la introducció d'una paraula clau i de considerar tants codis de Cèsar com lletres té l'alfabet. És a dir,

$$\begin{array}{cccccccc} \text{A} & \text{B} & \text{C} & \text{Ç} & \text{D} & \dots & \text{X} & \text{Y Z} \\ \text{B} & \text{C} & \text{Ç} & \text{D} & \text{E} & \dots & \text{Y} & \text{Z A} \\ \text{C} & \text{Ç} & \text{D} & \text{E} & \text{F} & \dots & \text{Z} & \text{A B} \\ \vdots & & & & & & & \\ \text{Z} & \text{A} & \text{B} & \text{C} & \text{Ç} & \dots & \text{W} & \text{X Y} \end{array}$$

Si volem codificar una frase com

BADA A CADA CAÇADA

i la paraula clau és CAZ, fem el següent:

BADA A CADA CAÇADA
CAZC A ZCAZ CAZCAZ

i la codificació de cada lletra s'obté buscant la lletra que hi ha sota de la lletra del missatge original que és a la fila que comença per la lletra corresponent de la paraula clau. A l'exemple l'encriptació de la B segons la fila que comença per C és Ç, i la de tot el missatge és:

ÇAÇC A BCDZ DACCDZ

La descodificació es fa usant la mateixa paraula clau i la taula en sentit contrari.

Aquest darrer mètode és molt més difícil de forçar (desencriptar) i és tan més complicat com més llarga és la paraula clau. L'encriptació perfecta, segons va provar C. Shannon, durant aquest segle s'aconsegueix quan la paraula clau té la mateixa longitud que el missatge que volem codificar i, a més, aquesta paraula clau està generada totalment a l'atzar. Això vol dir que la paraula clau es forma, per exemple, tirant un dau amb tantes cares com lletres té l'alfabet. Amb altres paraules, Shannon va provar que si s'utilitza una paraula clau aleatòria i de la mateixa mida que la frase que es vol codificar, la frase codificada és totalment aleatòria i per tant no ofereix cap informació sobre el missatge inicial, llevat que es conegui la clau de xifrat.

El problema d'aquest darrer sistema és evident: el receptor ha de conèixer una paraula clau tan llarga com el missatge, i aquesta paraula és difícil de canviar. Aquest problema és resol substituint la paraula clau totalment aleatòria pel que s'anomena una paraula pseudoaleatòria. Hi ha molts mètodes per a aconseguir paraules pseudoaleatòries; només n'explicarem un de molt senzill:

Considerem el número π:

$$3.141592653\,5897932384\,6264338327\,95\ldots$$

En els decimals de π no s'ha trobat cap regularitat i es poden considerar pseudoaleatoris. Una bona clau d'encriptació seria la següent: Prenem les deu primeres files de la taula d'encriptació deguda a Vigenère i les numerem del 0 al 9

0	A	B	C	Ç	D	...	Z
1	B	C	Ç	D	E	...	A
2	C	Ç	D	E	F	...	B
3	Ç	D	E	F	G	...	C
⋮							
9	I	J	K	L	M	...	H

Una clau d'encriptació de mida 100 pot ser els decimals de π des del 15 fins al 114.

Per tant, la frase i la clau d'encriptació serien

BADA A CADA CAÇADA
3 2 3 8 4 6 2 6 4 3 3 8 3 2 7

i s'encriptaria com DCGH...

Amb aquest últim mètode ja hem arribat als nostres dies. De fet, el sistema DES (Data Encription Standard) de IBM es basa en aquestes idees. L'única diferència important és que es treballa amb un altre alfabet més senzill: l'alfabet amb només dues "lletres": 0 i 1.

Un primer pas consisteix a traduir l'alfabet usual a zeros i uns; això s'anomena ASCII (American Standard Code of Information Interchange) i és l'usat pels ordinadors. En aquest codi:

A	és	01000001
B	és	01000010
C	és	01000011
Ç	és	10000000
D	és	01000100
⋮		

Aleshores, la paraula BADA en ASCII és

01000010 01000001 01000100 01000001.

Com a clau de codificació seguirem prenent els decimals de π entre 15 i 114, però convertits en una seqüència de cent zeros i uns amb la regla: una xifra parell dóna lloc a un 0, i una xifra senar dóna lloc a un 1. La clau corresponent a 323846... seria, per tant, 101000...

La taula de conversió en aquest cas segueix les mateixes regles i és molt senzilla (l'alfabet només té dues lletres). La podem escriure així:

El missatge té:	0	1
Clau 0	0	1
Clau 1	1	0

Per tant, la nostra paraula

01000010 01000001 01...

amb clau

10100000 01101011 10...

es codifica com

11100010 00101010 11...

Existeixen altres xifrats de substitució basats en idees diferents. A bastants d'aquests xifrats el coneixement de nombres primers[9] grans és molt important.

Xifrats de transposició

Parlarem ara d'una família de xifrats que no gaire cosa a veure amb l'anterior. En aquests xifrats, els valors de les lletres no varia mai. Pel que es coneix, també provenen de la Grècia clàssica. El mètode que explicarem s'atribueix a G. Cardano (segle *XVI*), encara que s'ha popularitzat per l'obra de Jules Verne *Mathias Sandorf*.

Els mètodes estan basats en una plantilla (nosaltres considerarem plantilles de 4×4) construïdes de la manera següent:

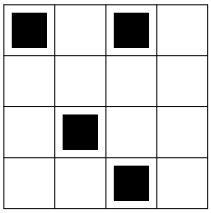
1	2	3	1
3	4	4	2
2	4	4	3
1	3	2	1

[9] L'estudi dels nombres primers ha estat una constant al llarg de la història de les matemàtiques. Potser en aquest llibre hi falta alguna secció dedicada a ells. No podem evitar comentar un parell de curiositats:

- No es coneix cap fórmula que ens doni sempre nombres primers. Euler va trobar el següent resultat sorprenent: la funció $f(n) = n^2+n+41$ dóna sempre un nombre primer quan l'avaluem a $n = 0, 1, 2, \ldots, 39$. Malauradament, $f(40) = 41^2$.
- Dos números primers $p < q$ es diuen bessons si $q - p = 2$. Per exemple, 3 i 5, 29 i 31, 149 i 151, 7949 i 7951, 104849 i 104851... són parells de nombres primers bessons. Avui en dia no se sap si hi ha o no infinits parells de primers bessons.

Observi's que el que s'ha fet és omplir el quadrat 2×2 de dalt a l'esquerra amb els números de l'1 al 4. Els altres 3 quadrats 2×2 s'omplen girant el quadrat gran 90, 180 i 270 graus en el sentit de les agulles del rellotge i posant als quadradets buits els nombres que van passant per sobre. Un cop fet això, es foraden quatre (o menys) quadradets respectant la regla següent: no es poden foradar quadrats amb nombres repetits.

Triem la plantilla següent



Aleshores el mètode d'encriptació funciona així: Donada una frase a encriptar com, per exemple,

LES MATEMÀTIQUES

es comença a escriure sobre un paper la frase usant només els forats de la plantilla. Un cop acabats els forats, es gira 90 graus la plantilla (en el sentit de les agulles del rellotge) i es continua escrivint i així dos cops més. El mètode seguit per a fer la plantilla ens garanteix que mai no haurem de sobreposar lletres. Obtenim

L	A	E	A
U	T	T	E
E	S	S	M
B	I	M	Q

on hem subratllat les quatre primeres lletres. Afegim una B o qualsevol altra lletra per a completar el quadrat de lletres. Per a acabar, les podem escriure com una frase

LAEA UT TEESS MBIMQ

Per a veure si el mètode és bo o no, el que hauria de passar és que encara que una persona l'interceptes no fos capaç de desxifrar-lo tot i sabent que s'ha usat el mètode de les plantilles. Per tant, la seguretat del mètode depèn del nombre diferent de plantilles que es puguin construir. En el cas 4×4 observeu que se'n poden construir 4^4 (triem un número 1 d'entre 4 per al primer forat, un 2 d'entre 4 per al segon, i així successivament). Si prenguéssim plantilles 6×6 o 8×8, el nombre possible de plantilles seria 4^9 o 4^{16}, i per tant el mètode seria molt més segur.

Tota l'exposició d'aquest tema esta basada en el llibre: *Códigos secretos,* de Andrea Sgamo, Ed. Pirámide SA, Madrid 1990.