Unidad 3: Criptografía

Definiciones

Criptografía (escritura oculta)

Rama de las matemáticas y de la informática que se ocupa de cifrar/descifrar información utilizando métodos y técnicas que permitan el intercambio de mensajes de manera que sólo puedan ser leídos por las personas a quienes van dirigidos.

Criptoanálisis

Estudio de los métodos que se utilizan para quebrar textos cifrados con objeto de recuperar la información original en ausencia de la clave.

Definiciones

Criptología

Ciencia que estudia las técnicas criptográficas y de criptoanálisis.

Cifra

Método o técnica que protege a un mensaje al aplicar un algoritmo criptográfico. Sin conocer una clave específica, no será posible descifrarlo o recuperarlo.

Esteganografía

Es la comunicación secreta lograda mediante la ocultación de la existencia de un mensaje.

Criptosistema

Tupla (E, D, M, K, C)

- M conjunto de textos en claro
- K conjunto de claves
- C conjunto de textos cifrados
- E conjunto de funciones de cifrado *e*: $M \times K \rightarrow C$
- D conjunto de funciones de descifrado *d*: $C \times K \rightarrow M$

Objetivo

El objetivo de la criptografía es mantener la información cifrada secreta.

<u>Atacante</u>

Alguien cuya meta es quebrar un criptosistema.

- Se asume que conoce el algoritmo pero no la clave.

Tipos de ataques

Fuerza bruta

Sólo texto cifrado

El atacante conoce solo el texto cifrado, su meta es encontrar el texto en claro y posiblemente la clave.

Texto en claro conocido

El atacante conoce solo el texto cifrado y el texto en claro, su meta es encontrar la clave.

Texto en claro elegido

El atacante usa varios textos en claro y obtiene los textos cifrados, su meta es encontrar la clave.

Tipos de ataques

Ataques matemáticos

- Basados en el análisis matemático de los algoritmos.

Ataques estadísticos

- Hacer suposiciones sobre la distribución de las letras (monogramas), los pares de letras (digramas), las ternas de letras (trigramas), etc.
- Examinar el texto cifrado y relacionar las propiedades con las suposiciones realizadas.

Criptografía clásica

Criptografía clásica

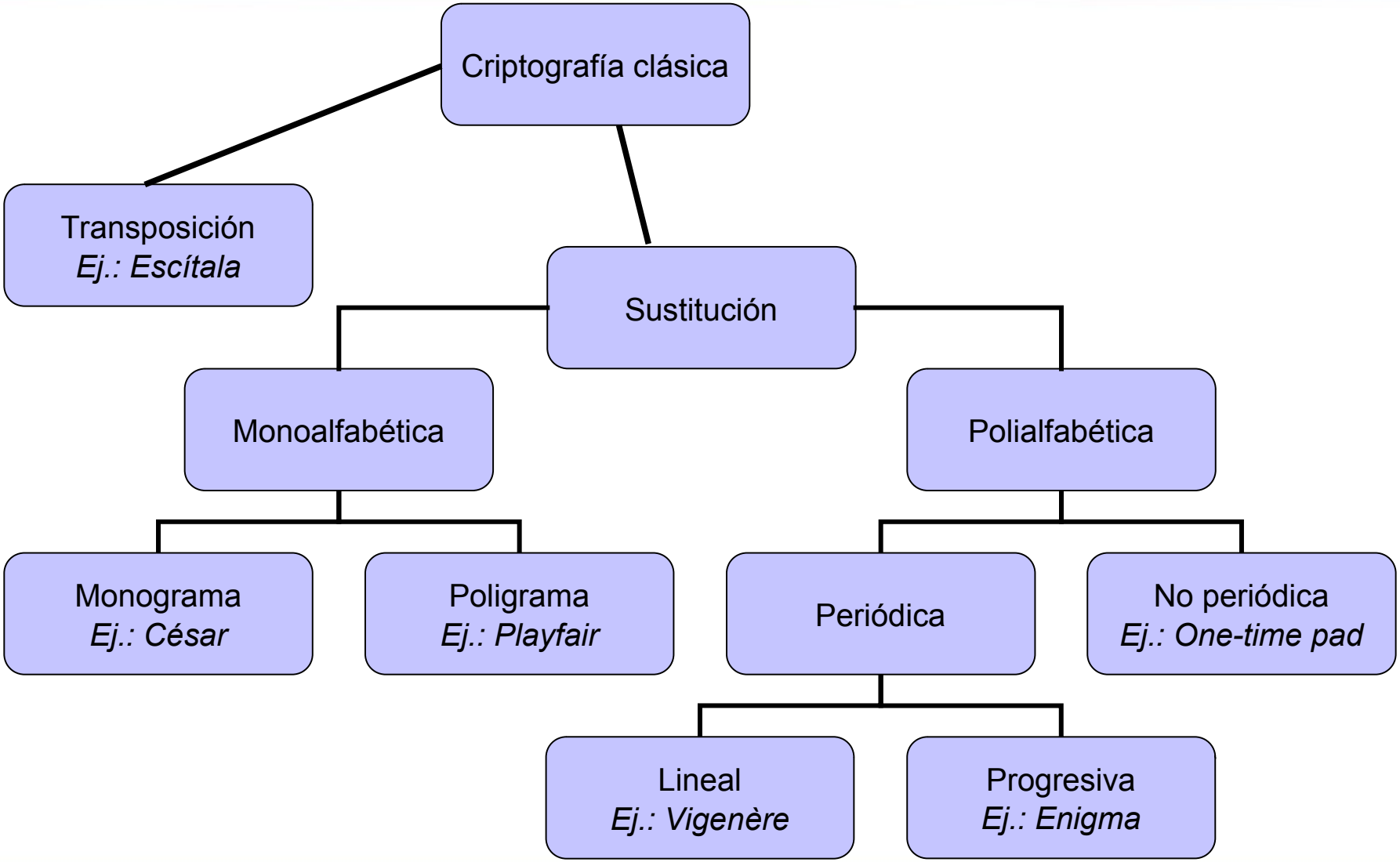
El emisor y el receptor comparten una clave

- Las claves pueden ser la misma o fácilmente derivables una de la otra.
- Denominado *criptografía simétrica.*

Tipos

- Cifra por transposición.
- Cifra por sustitución.
- Cifra por combinación de los anteriores (Producto).

Criptografía clásica



Transposición

Consiste en reorganizar los caracteres del texto en claro para producir el texto cifrado.

En muchos casos se sigue una pauta geométrica: se escribe el mensaje en un rectángulo por filas y se lee por columnas, etc.

Por ejemplo:

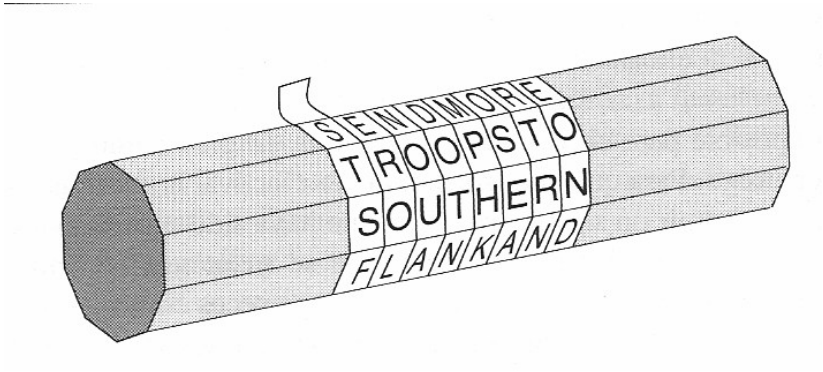
- Texto en plano: `HELLO WORLD`
- Reorganizarlo de la siguiente manera

  `HLOOL`

  `ELWRD`
- Texto cifrado: `HLOOL ELWRD`

Cifra de la escítala

- **Utilizada en Grecia en el Siglo V a.c.**
- **La escítala es una vara de madera sobre la que se enrosca una tira de cuero. El emisor escribe el mensaje a lo largo de la vara y luego desenrosca la tira.**
- **La tira solía ocultarse como cinturón.**
- **Para recuperar el mensaje el receptor necesita una vara igual a la usada para crearlo.**



sendmoretroopstosouthernflankand

STSFEROLNOUADOTNMPHKOSEARTRNEOND

Sustitución

Consiste en cambiar caracteres del texto en claro para producir el texto cifrado.

Es decir, cada letra del mensaje se sustituye por otra según una determinada permutación del alfabeto.

Por ejemplo:

- Texto en plano: `HELLO WORLD`
- Cambiar cada letra por la tercer letra que le sigue (A por D, B por E, C por F)
- Texto cifrado: `KHOOR ZRUOG`

Tipos de sustitución

Sustitución monoalfabética

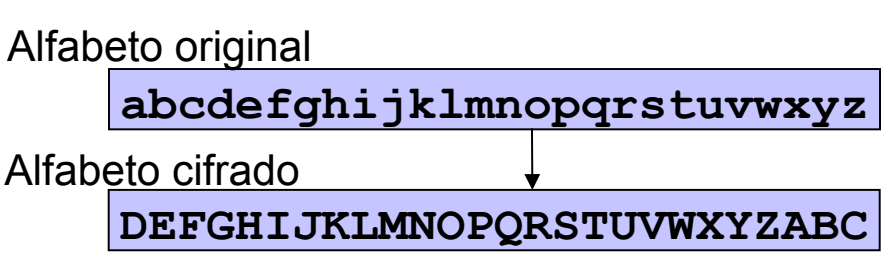
Cuando a una misma letra del mensaje le corresponde siempre la misma letra del texto cifrado.

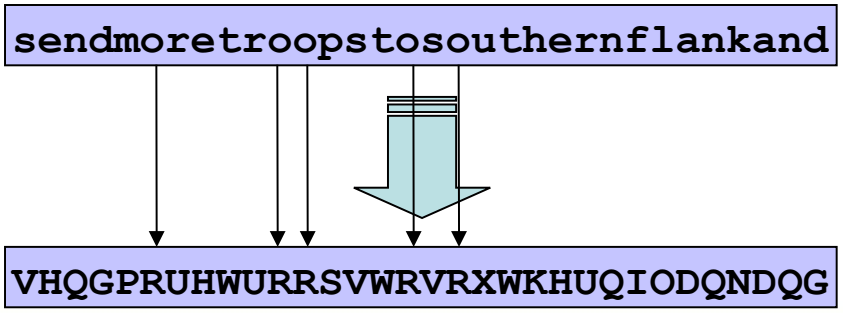
Sustitución polialfabética

Cuando a una misma letra del mensaje le corresponden distintas letras del mensaje cifrado.

Cifra del César

- **Es una sustitución monoalfabética monograma.**
- **Se dice que era utilizada por Julio César para la comunicación de mensajes a sus tropas.**
- **Consiste en reemplazar cada letra del mensaje original por la letra que se encuentra 3 lugares adelante en el alfabeto,**
  por ejemplo a→D, b→E, … z→C.





Cifra del César

- **M = { secuencia de letras del alfabeto }**
- **K = { i | i es un número entero y $0 \leq i \leq 26$ }**
- **E = { E_k | $k \in$ K y para todas las letras m,**

  $E_k(m) = (m + k) \bmod 27$ }
- **D = { D_k | $k \in$ K para todas las letras c,**

  $D_k(c) = (27 + c - k) \bmod 27$ }
- **C = M**

Análisis de frecuencia

- Se basa en el análisis de la frecuencia de aparición de los símbolos del texto cifrado y su intento de correlación con los símbolos del lenguaje en el cual está escrito el mensaje.
- Se buscan los caracteres más frecuentes en el criptograma y se los asocia a las letras de mayor aparición en el idioma original.
- Se prueban distintas alternativas hasta alcanzar un texto coherente.
- Es una herramienta criptoanalítica de base.

Análisis de frecuencias

a	10.60	j	0.25	r	0.74
b	1.16	k	0.11	s	8.47
c	4.85	l	4.42	t	5,40
d	5.87	m	3.11	u	4,34
e	13.11	n	7.14	v	0.82
f	1.13	ñ	0.10	w	0.12
g	1.40	o	8.23	x	0.15
h	0.60	p	2.71	y	0.79
i	7.16	q	0.74	z	0.26

Tabla de frecuencias del español

Cifra de Vigenère

En lugar de una letra ahora utilizo una frase.
Por ejemplo:

- Texto en plano: `THE BOY HAS THE BALL`
- Clave: `VIG`
- Cifrar utilizando la cifra de César para cada letra:

```
key      VIGVIGVIGVIGVIGV
plain    THEBOYHASTHEBALL
cipher   OPKWWECIYOPKWIRG
```

El tableau

Tabla que tiene las letras de la clave en la parte superior (columnas), y las letras de texto en claro a la izquierda (filas).

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

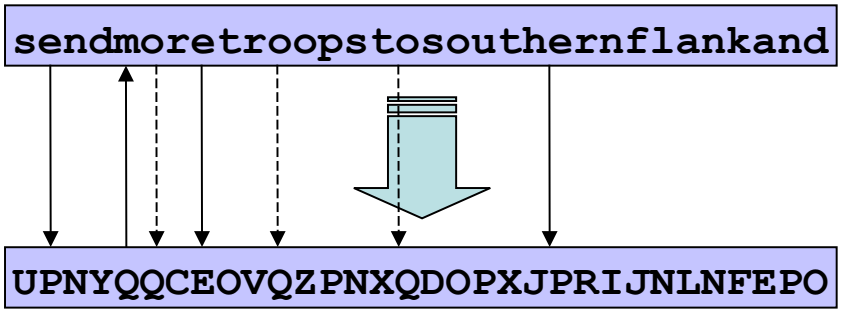
Ejemplo:

- Clave V, letra T: seguir por la columna V hasta la fila T, se obtiene “O”
- Clave I, letra H: seguir por la columna I hasta la fila H, se obtiene “P”

Cifra de Vigenère

Sustitución polialfabética periódica lineal, con período igual a la longitud de la clave.

Fue denominada durante muchos años como “la cifra indescifrable”.



Ataque de Kasiski (1863)

- La cifra de Vigenère puede ser atacada con éxito mediante un análisis de frecuencia
- La técnica consiste en:
  - Buscar cadenas repetidas.
  - Buscar el período de la clave obteniendo el *MCD* (máximo común divisor) entre las posiciones de todas las cadenas repetidas.
  - Descomponer el problema en N sistemas. monoalfabéticos (donde N es el tamaño de la clave)
  - Abordar cada sistema monoalfabético por medio del análisis de frecuencias.

Ataque de Kasiski

En el texto cifrado ocurren repeticiones cuando los caracteres de la clave aparecen sobre los mismos caracteres del texto en claro.

Por ejemplo:

```
key     VIGVIGVIGVIGVIGV
plain   THEBOYHASTHEBALL
cipher  OPKWWECIYOPKWIRG
```

La clave y el texto en claro se alinean sobre las repeticiones. La distancia entre las repeticiones es 9 y por lo tanto el periodo es factor de 9 (1, 3 o 9).

Ejemplo

Queremos quebrar este texto cifrado:

```
ADQYS MIUSB OXKKT MIBHK IZOOO
EQOOG IFBAG KAUMF VVTAA CIDTW
MOCIO EQOOG BMBFV ZGGWP CIEKQ
HSNEW VECNE DLAAV RWKXS VNSVP
HCEUT QOIOF MEGJS WTPCH AJMOC
HIUIX
```

Tabla de repeticiones

Letras	*Inicio*	*Fin*	*Distancia*	*Factores*
MI	5	15	10	2, 5
OO	22	27	5	5
OEQOOG	24	54	30	2, 3, 5
FV	39	63	24	2, 2, 2, 3
AA	43	87	44	2, 2, 11
MOC	50	122	72	2, 2, 2, 3, 3
QO	56	105	49	7, 7
PC	69	117	48	2, 2, 2, 2, 3
NE	77	83	6	2, 3
SV	94	97	3	3
CH	118	124	6	2, 3

Estimar el período

- **OEQOOG**
  - Es una repetición muy larga
  - El período puede ser 1, 2, 3, 5, 6, 10, 15, o 30
- **7 de 10 repeticiones tiene a 2 en sus factores**
- **6 de 10 repeticiones tiene a 3 en sus factores**
- **Comenzamos entonces con un período $2 \times 3 = 6$**

Índice de coincidencia

Es la probabilidad de que dos letras de un texto cifrado elegidas al azar sean la misma.

$$IC = [n\,(n-1)]^{-1} \sum_{0 \le i \le 25} [F_i\,(F_i - 1)]$$

donde n es la longitud del texto cifrado y F_i la cantidad de veces que aparece la letra i en dicho texto.

Se encuentra tabulado para diferentes períodos:

1	0.066	30.047	50.044
2	0.052	40.045	100.041
más de 10		0.038	

Obtener los alfabetos

alfabeto 1: `AIKHOIATTOBGEEERNEOSAI`

alfabeto 2: `DUKKEFUAWEMGKWDWSUFWJU`

alfabeto 3: `QSTIQBMAMQBWQVLKVTMTMI`

alfabeto 4: `YBMZOAFCOOFPHEAXPQEPOX`

alfabeto 5: `SOIOOGVICOVCSVASHOGCC`

alfabeto 6: `MXBOGKVDIGZINNVVCIJHH`

ICs (#1, 0.069; #2, 0.078; #3, 0.078; #4, 0.056; #5, 0.124; #6, 0.043) indican que los alfabetos tienen período 1, excepto #4 y #6.

Análisis de frecuencias


```
   ABCDEFGHIJKLMNOPQRSTUVWXYZ
1  31004011301001300112000000
2  10022210013010000010404000
3  12000000201140004013021000
4  21102201000010431000000211
5  10500021200000500030020000
6  01110022311012100000030101
```

Las frecuencias del idioma son (H high, M medium, L low):

HMMMHMMHHMMMMHHMLHHHMLLLLL

Descifrado

Se analiza cada alfabeto para ver cual es su desplazamiento.

- El primero queda igual
- El tercero tiene la I cambiada por A
- El sexto tiene la V cambiada por A
- Se realizan las sustituciones (negrita es lo cifrado)

AD**I**YS **RI**U**K**B O**CK**K**L** MI**GH**K **A**ZO**TO** E**I**OO**L**
IF**T**AG **PA**U**E**F V**AT**A**S** C**I**I**T**W **E**OC**NO** E**I**OO**L**
BM**T**FV **EG**G**O**P C**NE**K**I** HS**SE**W **N**EC**SE** D**D**AA**A**
RW**C**XS **AN**S**N**P H**HE**U**L** QO**NO**F **E**EG**OS** W**L**PC**M**
AJ**E**OC **MI**U**A**X

Descifrado

El análisis continua buscando en el texto pistas que sugieran palabras conocidas para de esta manera determinar los desplazamientos de los alfabetos restantes.

Luego de terminar el proceso tendremos:

Clave: `ASIMOV`

Texto en claro:

```
ALIME RICKP ACKSL AUGHS ANATO MICAL INTOS
PACET HATIS QUITE ECONO MICAL BUTTH EGOOD
ONESI VESEE NSOSE LDOMA RECLE ANAND THECL
EANON ESSOS ELDOM ARECO MICAL
```

One-time pad (cuaderno de uso único)

- **Es una variante de Vigenère con una clave aleatoria tan larga como el mensaje en claro.**
- **Es un sistema criptográfico "perfectamente seguro".**

  El texto cifrado `DXQR`, puede corresponder al texto en claro `DOIT` (cifrado con la clave `AJIY`) y al texto en claro `DONT` (cifrado con la clave `AJDY`) y a cualquier otra combinación de 4 letras.

- **Las claves deben ser aleatorias, de no serlo se puede atacar tratando de regenerar la clave.**
- **Las claves se deben usar una sola vez.**

Criptografía moderna

Definiciones

Principios de Kerckhoffs (1883)

- El sistema debe ser en la práctica imposible de criptoanalizar.
- La seguridad de un sistema criptográfico debe depender **sólo** de que la clave sea secreta y **no** de que el algoritmo de cifrado sea secreto
- Método de elección de claves fácil de recordar.
- Transmisión del texto cifrado por telégrafo.
- La máquina de cifrar debe ser portable.
- No debe existir una larga lista de reglas de uso.

Definiciones

En 1948 Claude Shannon publica el artículo *A Mathematical Theory of Communication*, sentando las bases de la Teoría de la Información.

Información

- Es el conjunto de datos o mensajes inteligibles creados con un lenguaje de representación.
- Ante varios mensajes posibles, aquel que tenga una menor probabilidad de aparición será el que contenga una mayor cantidad de información.

Definiciones

Seguridad perfecta

Un sistema criptográfico es perfectamente seguro si el texto cifrado no da ninguna información adicional del texto plano. Es decir que dado un texto cifrado C, cualquier posible texto plano es igualmente probable con respecto a C.

Teoremas

- En todo sistema perfectamente seguro, la longitud de las claves es mayor o igual que la de los mensajes.
- Existen sistemas perfectamente seguros. (por ejemplo: one-time pad)

Definiciones

Sistema incondicionalmente seguro

Cuando es seguro frente a ataques con capacidad de cálculo ilimitada.

Sistema computacionalmente seguro

Cuando es seguro frente a ataques con capacidad de cálculo limitada.

Definiciones

Para mejorar las operaciones de cifra Shannon propone dos técnicas:

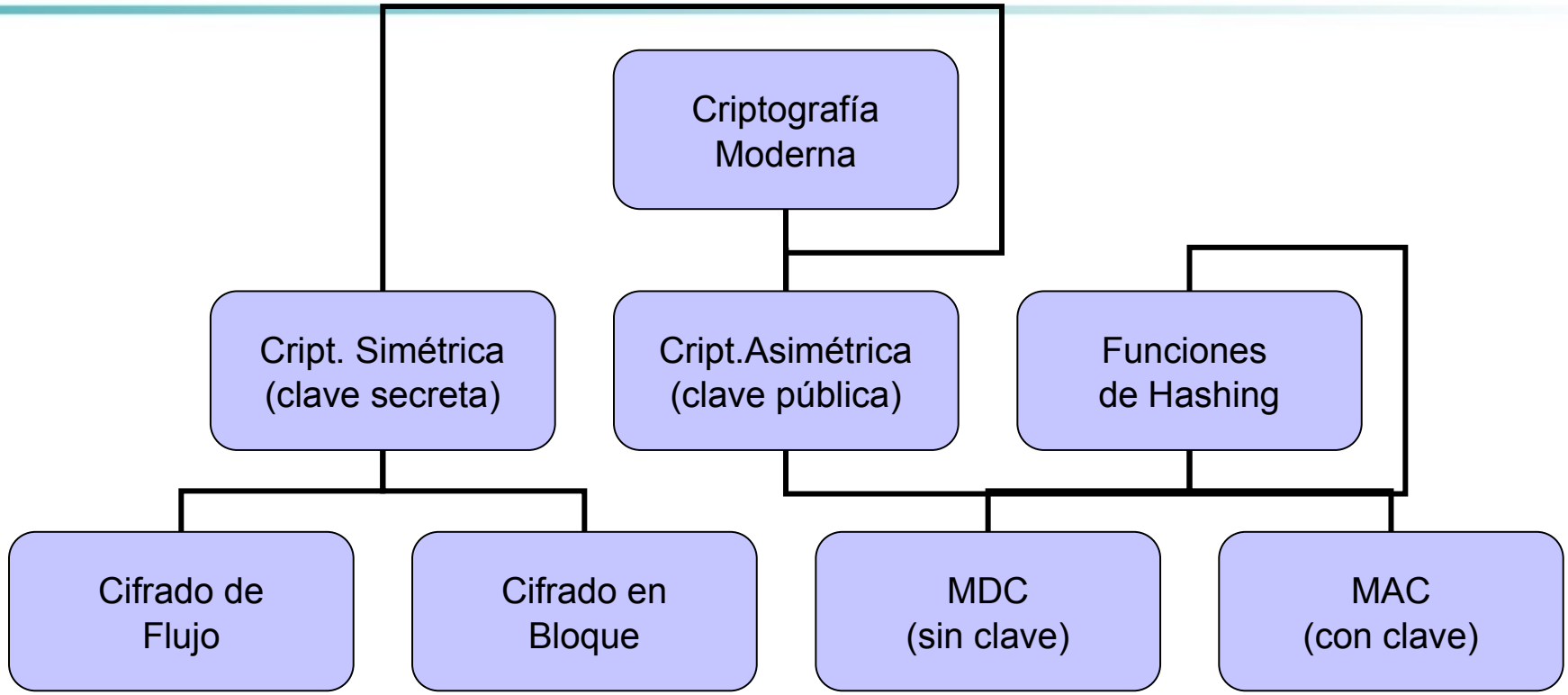
<u>Difusión</u>

Es la transformación del texto claro con el objeto de dispersar las propiedades estadísticas del lenguaje sobre el criptograma. Se logra con *Transposiciones*.

<u>Confusión</u>

Es la transformación del texto claro con el objeto de mezclar los elementos de éste, aumentando la complejidad de la dependencia funcional entre la clave y el criptograma. Se obtiene mediante *Sustituciones*.

Criptografía moderna



Criptografía simétrica

- **También conocidos como criptografía de c*lave secreta* o c*lave privada.***
- **La clave utilizada en la operación de cifrado es la misma que se utilizada para el descifrado.**
- **Existen dos mecanismos de operación básicos:**
  - Cifrado de flujo.
  - Cifrado en bloques.

Cifrado de flujo y cifrado en bloque

Sea *E* una función de cifrado

- $E_k(b)$ cifrado del mensaje b con la clave k
- $m = b_1 b_2 \ldots$, donde cada b_i es de longitud fija

Cifrado en bloque

- $E_k(m) = E_k(b_1)E_k(b_2) \ldots$

Cifrado de flujo

- $k = k_1 k_2 \ldots$
- $E_k(m) = E_{k1}(b_1)E_{k2}(b_2) \ldots$
- Si $k_1 k_2 \ldots$ se repite, el cifrador es *periódico* y la longitud de su periodo es un ciclo de $k_1 k_2 \ldots$

Cifradores de flujo

Usan los siguientes conceptos:

- El espacio de las claves es igual o mayor que el espacio de los mensajes.
- Las claves son aleatorias.
- La secuencia de clave se usa sólo una vez.

La solución ideal no es práctica.

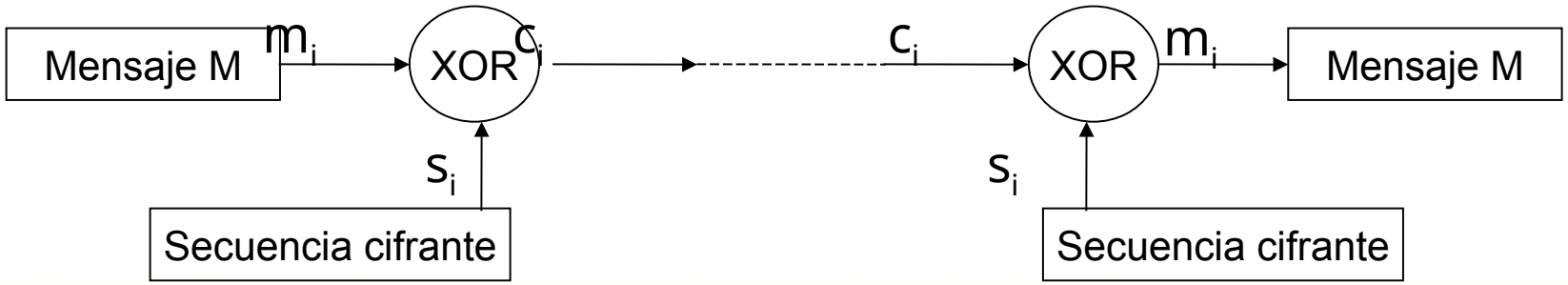
Cifradores de flujo

Se utilizan generadores pseudoaleatorios con un algoritmo determinístico a partir de una semilla de n bits, pudiendo generar secuencias con periodos de 2^n bits.

Lo que se transmite es solo la semilla.

Modo de operación

- **El mensaje en claro se lee y transmite bit a bit**
- **Se realiza una operación de cifra, normalmente la función XOR, entre los bits del mensaje en claro y los de la secuencia cifrante.**
- **La secuencia cifrante debe cumplir:**
  - Tener un período muy alto.
  - Tener propiedades pseudoaleatorias.



Modo de operación

Existen 2 mecanismos de operación de las secuencias cifrantes:

- Sincrónicos

  El emisor y el receptor deben sincronizarse previamente a la transmisión.

  La perdida de 1 bit en el flujo de datos puede inutilizar el resto de la transmisión.

- Auto-sincrónicos

  Se utiliza parte de la información del texto cifrado para renovar la clave de la secuencia cifrante.

Cifradores en bloque

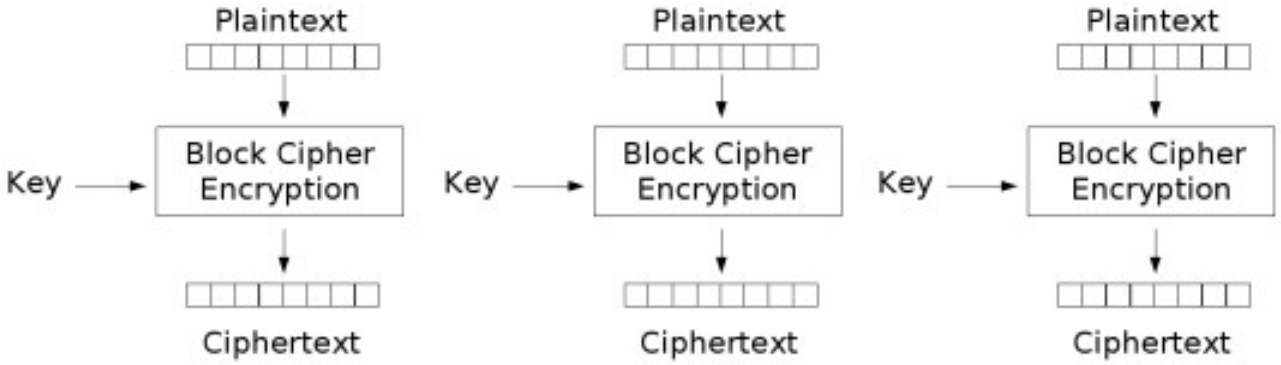
El mensaje se divide en bloques de longitud fija (8, 16, … bytes) y luego se aplica el algoritmo de cifrado a cada bloque en forma independiente con la misma clave

Existen distintos modos de operación dependiendo de cómo se mezcla la clave con el texto en claro. Algunos ejemplos iniciales:

- EBC: Electronic CodeBook
- CBC: Cipher Block Chaining
- CTR: Counter Mode

ECB: Electronic CodeBook

- **El texto se divide en bloques y cada bloque es cifrado en forma independiente utilizando la clave.**



Electronic Codebook (ECB) mode encryption

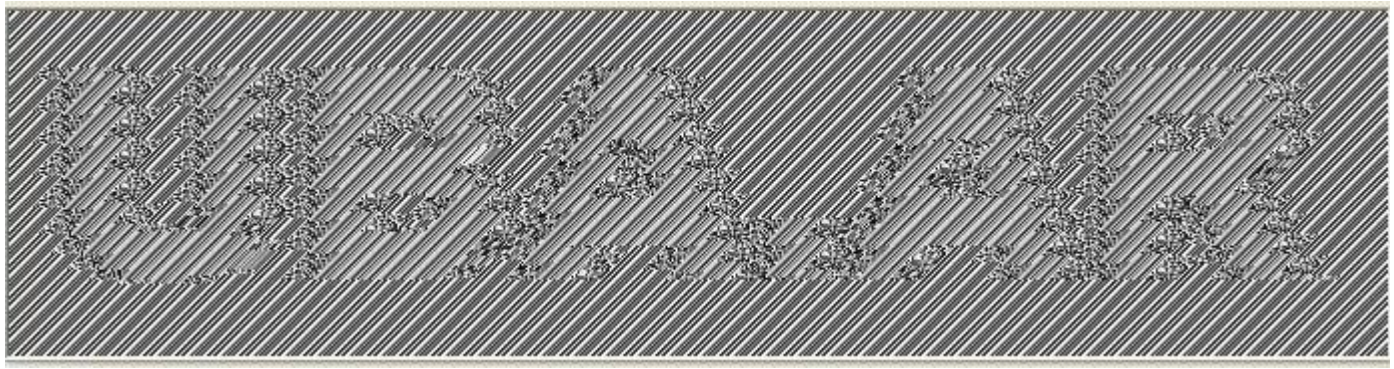
Ejemplo problemas ECB

- **Tengo una imagen en formato raw:**

- Y la cifro utilizando AES-128, con ECB:

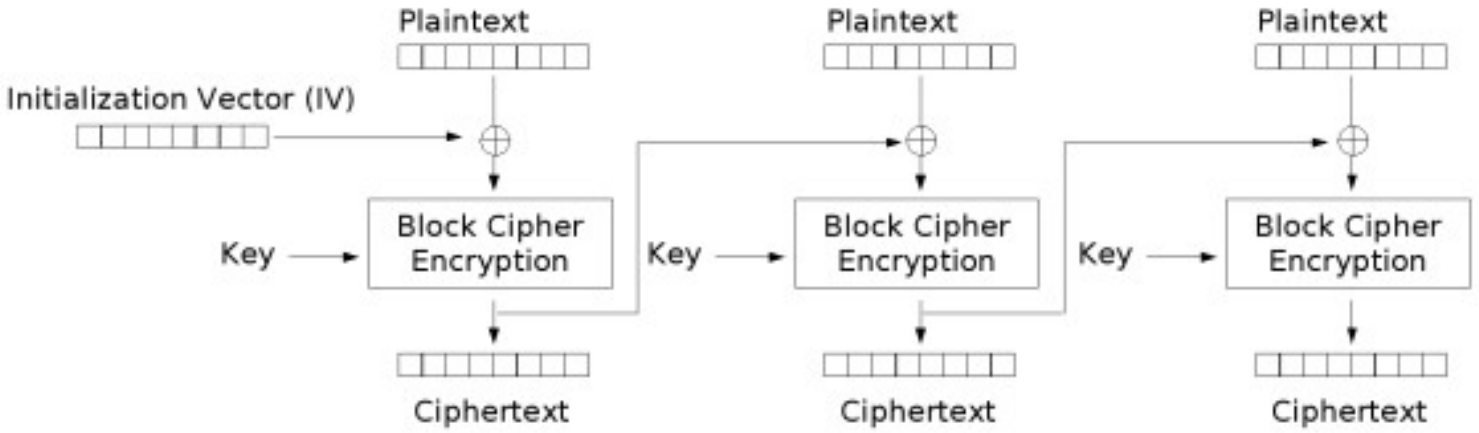
Cont. Problemas ECB

- **Obtengo una nueva imagen raw, y puedo entender su contenido, pese a estar encriptada!**



CBC: Cipher-Block Chaining

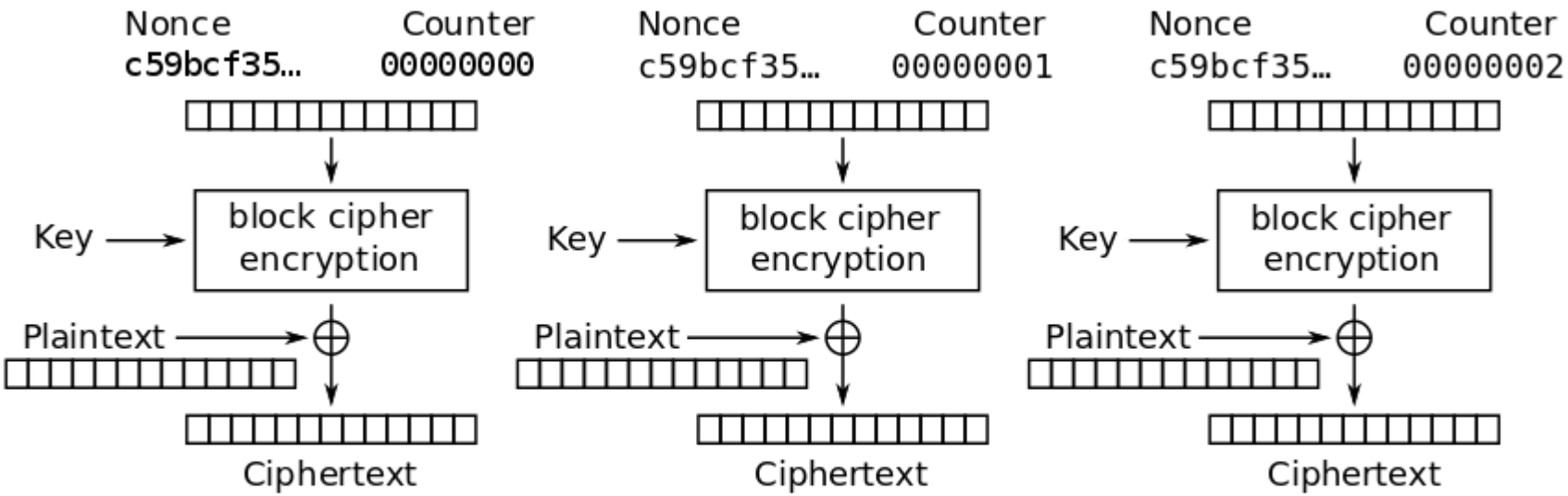
- **El texto se divide en bloques y cada bloque es mezclado con la cifra del bloque previo, luego es cifrado utilizando la clave.**



Cipher Block Chaining (CBC) mode encryption

CTR: Counter mode

- **Se usa un nounce + un contador. Se puede paralelizar.**



Counter (CTR) mode encryption

Padding

Al dividir el texto original en bloques de longitud fija, algunos modos de cifrado requieren que se rellene el último bloque (padding) antes de realizar la operación.

Este texto de relleno debe ser quitado durante la operación de descifrado.

Por ejemplo:

- Se agrega 1 bit en 1 seguido de ceros hasta completar el bloque.

  Si el texto terminaba exacto se agrega un bloque completo.

Padding en estándar PKCS#5

- **Estándar PKCS#5: El último bloque se completa con N bytes con valor N, si es múltiplo se completa con un bloque completo de padding. Lo mismo usa PKCS#7**

- **Gráficamente:**

Primer Bloque							
1	2	3	4	5	6	7	8
A	A	A	A	A	A	A	A
A	A	A	A	A	A	A	A
...							
A	A	A	A	A	A	A	A
A	A	A	A	A	A	A	A

Segundo Bloque							
1	2	3	4	5	6	7	8
B	B	B	B	B	B	B	0x01
B	B	B	B	B	B	0x02	0x02
...							
B	0x07	0x07	0x07	0x07	0x07	0x07	0x07
0x08	0x08	0x08	0x08	0x08	0x08	0x08	0x08