Aula virtual de criptografía aplicada

Diapositivas utilizadas en las clases grabadas de Class4crypt

Módulo 6 Algoritmos de criptografía clásica

Dr. Jorge Ramió Aguirre © 2022

Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0)

https://www.youtube.com/user/jorgeramio

Dr. Jorge Ramió Aguirre

El ingenio es intrínseco al ser humano, solo hay que darle una oportunidad para que se manifieste.

https://www.criptored.es/cvJorge/index.html

Class4crypt

Tu aula virtual de criptografía aplicada

Módulo 1. Principios básicos de la seguridad
Módulo 2. Matemáticas discretas en la criptografía
Módulo 3. Complejidad algorítmica en la criptografía
Módulo 4. Teoría de la información en la criptografía
Módulo 5. Fundamentos de la criptografía
→ Módulo 6. Algoritmos de criptografía clásica
Módulo 7. Funciones hash
Módulo 8. Criptografía simétrica en bloque
Módulo 9. Criptografía simétrica en flujo
Módulo 10. Criptografía asimétrica

© jorgeramio 2022

Class4crypt

Módulo 6. Algoritmos de criptografía clásica

6.1 Cifrado por permutación o transposición (parte 1 y parte 2)

6.2 Criptoanálisis a la cifra por permutación

6.3 Cifrado por sustitución

6.4 Criptoanálisis a la sustitución monoalfabética

6.5 Algoritmo de Vigenère

6.6 Criptoanálisis a la cifra de Vigenère por el método Kasiski

6.7 Cifrado digrámico de Playfair

6.8 Cifrado poligrámico con matrices de Hill

6.9 Criptoanálisis a la cifra de Hill por Gauss Jordan

Lista de reproducción del módulo 6 en el canal Class4crypt
https://www.youtube.com/playlist?list=PLq6etZPDh0kt4kk5CKDeHtNSehillvSgB

Class4crypt c4c6.1a

Módulo 6. Algoritmos de criptografía clásica

Lección 6.1a. Cifrado por permutación o transposición parte 1

6.1a.1. Técnica de permutación para lograr la difusión

6.1a.2. Cifrado por escítala

6.1a.3. Cifrado por filas y columnas

Class4crypt c4c6.1a Cifrado por permutación o transposición parte 1
https://www.youtube.com/watch?v=aY6D0pcWxxk

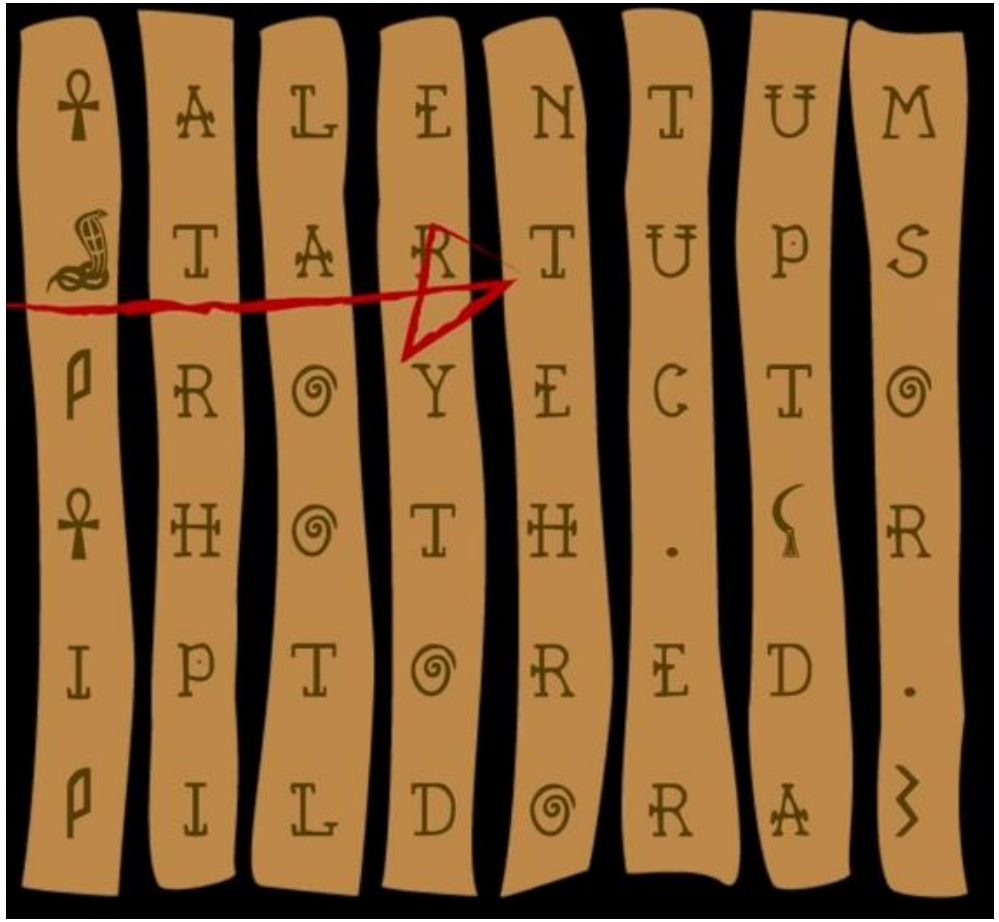
Permutaciones para lograr la difusión

- Esta técnica de cifra consiste en que las letras del texto en claro se redistribuyen dentro del criptograma sin sustituirlas, siguiendo unas reglas o patrón definidos
- Con esto se logra la difusión, es decir, difundir las características del texto en claro (del lenguaje) en todo el texto cifrado
- Ocultamos así la relación entre el texto en claro y el criptograma
- Aunque sería fácilmente detectable, podríamos lograr este efecto de difusión simplemente escribiendo un texto de forma diferente a como se suele hacer; por ejemplo, desde la derecha del papel hacia la izquierda y/o desde abajo hacia arriba

La escítala de Lacedemonia

- En Lacedemonia o Laconia, región de Grecia con capital Esparta, se inventa en el siglo V a.C. el primer sistema de cifra de la historia
- La escítala consiste en un bastón de un cierto diámetro y largo
  - Sobre el bastón se enrolla una cinta
  - Se va escribiendo el texto en esa cinta en forma longitudinal
  - Cuando se termina esa línea debido al largo del bastón, este se gira levemente y se sigue escribiendo en la línea o fila inferior
  - Si después de varios giros sigue habiendo espacios en la cinta sin texto, se rellenan esos espacios con caracteres aleatorios
  - Se desenrolla la cinta y ésta muestra el criptograma obtenido

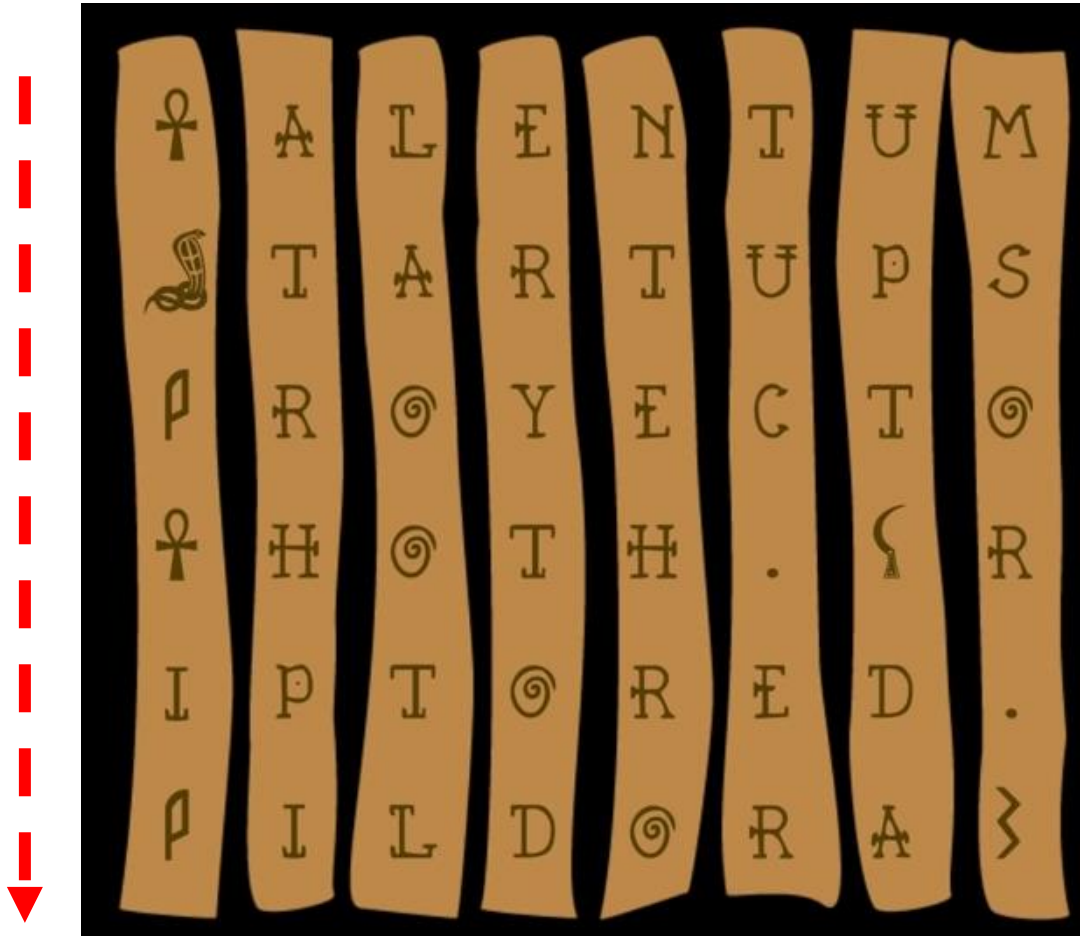
Cifrando con la escítala



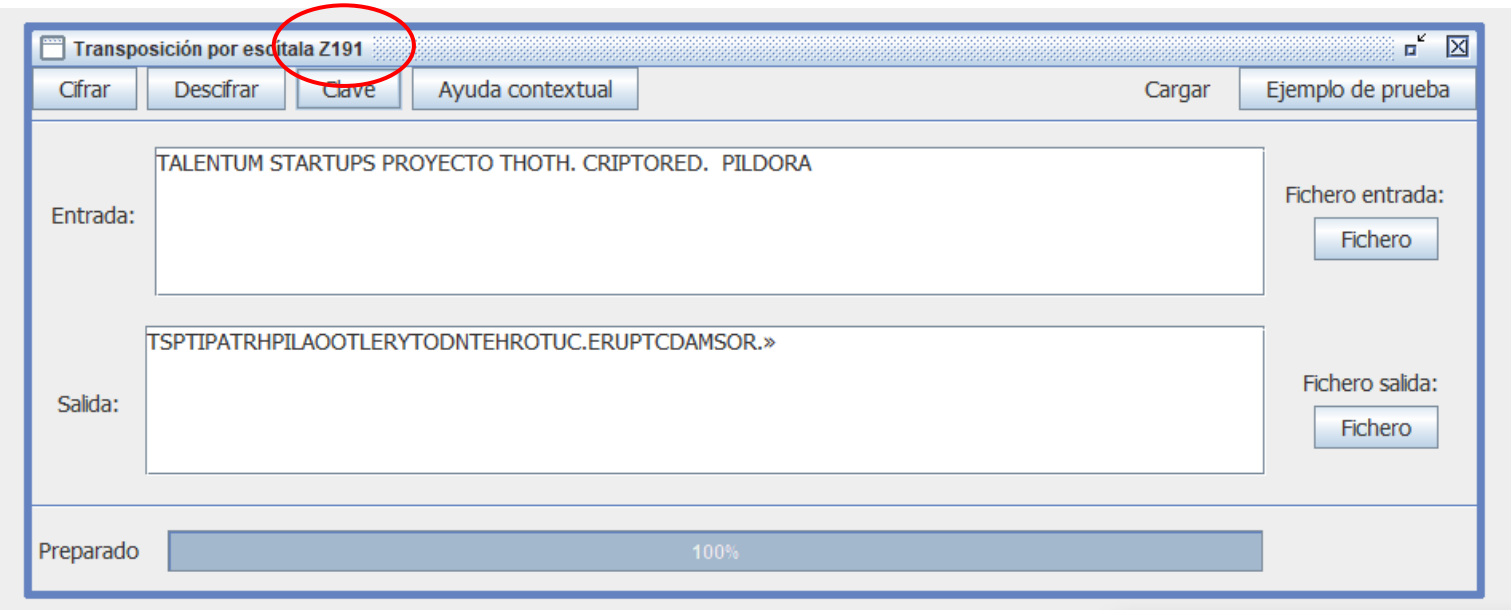
- Datos del bastón: vueltas 6, longitud 8
- M = TALENTUM STARTUPS PROYECTO THOTH. CRIPTORED. PILDORA
- C = TSPTIP ATRHPI LAOOTL ERYTOD NTEHRO TUC.ER UPTCDA MSOR.3
- Se ha usado como relleno el signo (3)

Descifrando con la escítala

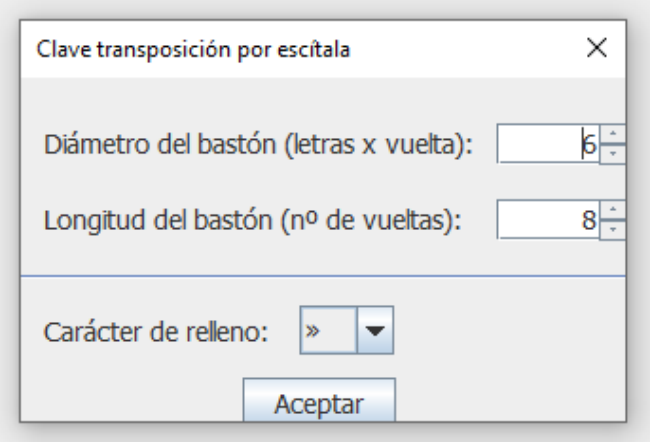
- C = TSPTIP ATRHPI LAOOTL ERYTOD NTEHRO TUC.ER UPTCDA MSOR.3
- Clave: vueltas 6, longitud 8
- Vamos escribiendo por columnas de arriba hacia abajo los 48 caracteres del criptograma y cambiamos de columna cada 6 caracteres (que es el diámetro del bastón)
- Terminada la operación, se lee el texto en claro por filas de arriba hacia abajo
- M = TALENTUM STARTUPS PROYECTO THOTH. CRIPTORED. PILDORA



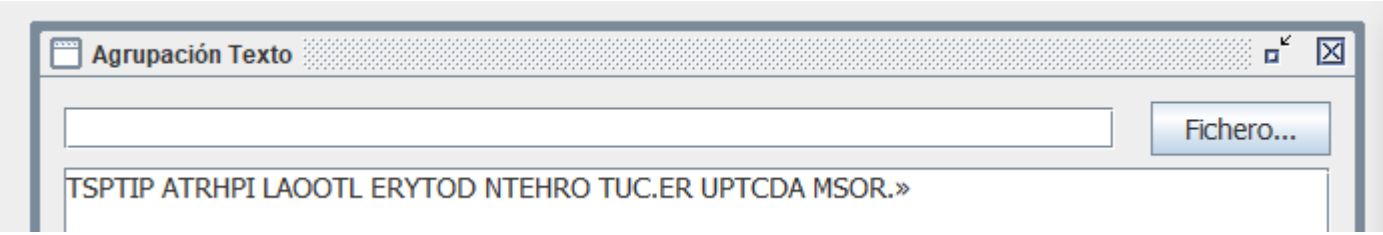
Práctica de escítala con Criptoclásicos



El texto tiene 47 caracteres sin contar los espacios en blanco



Agrupando el criptograma en bloques de seis caracteres



Cifrando por filas

- El texto en claro se escribe en vertical, de arriba hacia abajo, en las columnas de una matriz de n filas por m columnas
- Cada x filas, siendo x el número de la clave, se continúa escribiendo el texto que falta en la columna de la derecha
- El criptograma es el texto de la lectura de las filas de la matriz

- M = QUE LA FUERZA TE ACOMPAÑE
- Clave NF = 4, Relleno = X

Q	A	R	E	M	E
U	F	Z	A	P	X
E	U	A	C	A	X
L	E	T	O	Ñ	X

- C = QAREM EUFZA PXEUA CAXLE TOÑX

Descifrando por filas

- Conocida la longitud del texto cifrado L y el número NF de filas, se calcula el número de columnas NC = L/NF
- Se procede ahora a escribir el criptograma por filas hasta la columna NC, tras lo cual se sigue en la fila de abajo
- En las columnas de derecha a izquierda está el texto en claro

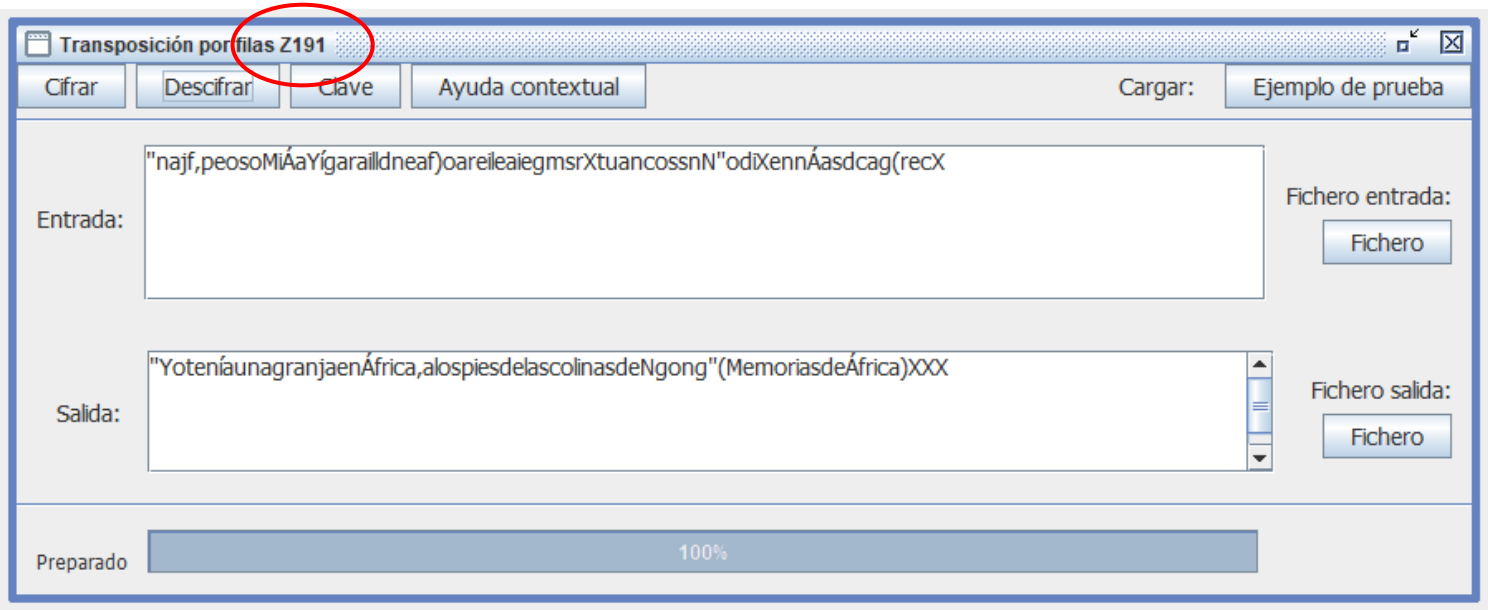
- C = EANEE SNSEO RXOIS MTXCO VUOX
- L = 24, NF = 4, NC = 24/4 = 6

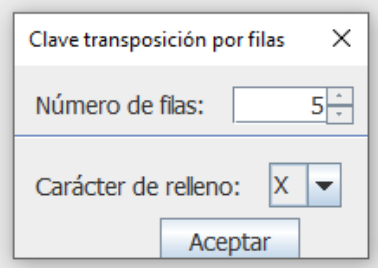
E	A	N	E	E	S
N	S	E	O	R	X
O	I	S	M	T	X
C	O	V	U	O	X

- M = EN OCASIONES VEO MUERTOS

Práctica de filas con Criptoclásicos

- C = "najf,peos oMiÁaYígar ailldneaf) oareileaie gmsrXtuanc ossnN"odiX ennÁasdcag (recX





"	n	a	j	f	,	p	e	o	s	o	M	i	Á	a
Y	i	g	a	r	a	i	l	l	d	n	e	a	f	)
o	a	r	e	i	l	e	a	i	e	g	m	s	r	X
t	u	a	n	c	o	s	s	n	N	"	o	d	i	X
e	n	n	Á	a	s	d	c	a	g	(	r	e	c	X

- M = "Yo tenía una granja en África, a los pies de las colinas de Ngong“ (Memorias de África)

Cifrando por columnas

- Se escribe en horizontal el texto en claro de izquierda a derecha, en las filas de una matriz de n filas por m columnas
- Cada x columnas, siendo x el número clave, el texto que falta se sigue escribiendo en la fila de abajo y siguientes
- El criptograma es el texto que se lee en las columnas de la matriz

- M = QUE LA FUERZA TE ACOMPAÑE
- Clave NC = 6, Relleno = X

Q	U	E	L	A	F
U	E	R	Z	A	T
E	A	C	O	M	P
A	Ñ	E	X	X	X

- C = QUEAU EAÑER CELZO XAAMX FTPX

Descifrando por columnas

- Conocida la longitud del texto cifrado L y el número NC de columnas, se calcula el número de filas NF = L/NC
- Se procede a escribir el criptograma por columnas hasta la fila NF, tras lo cual se sigue en la columna derecha
- En las filas de arriba abajo está el texto en claro

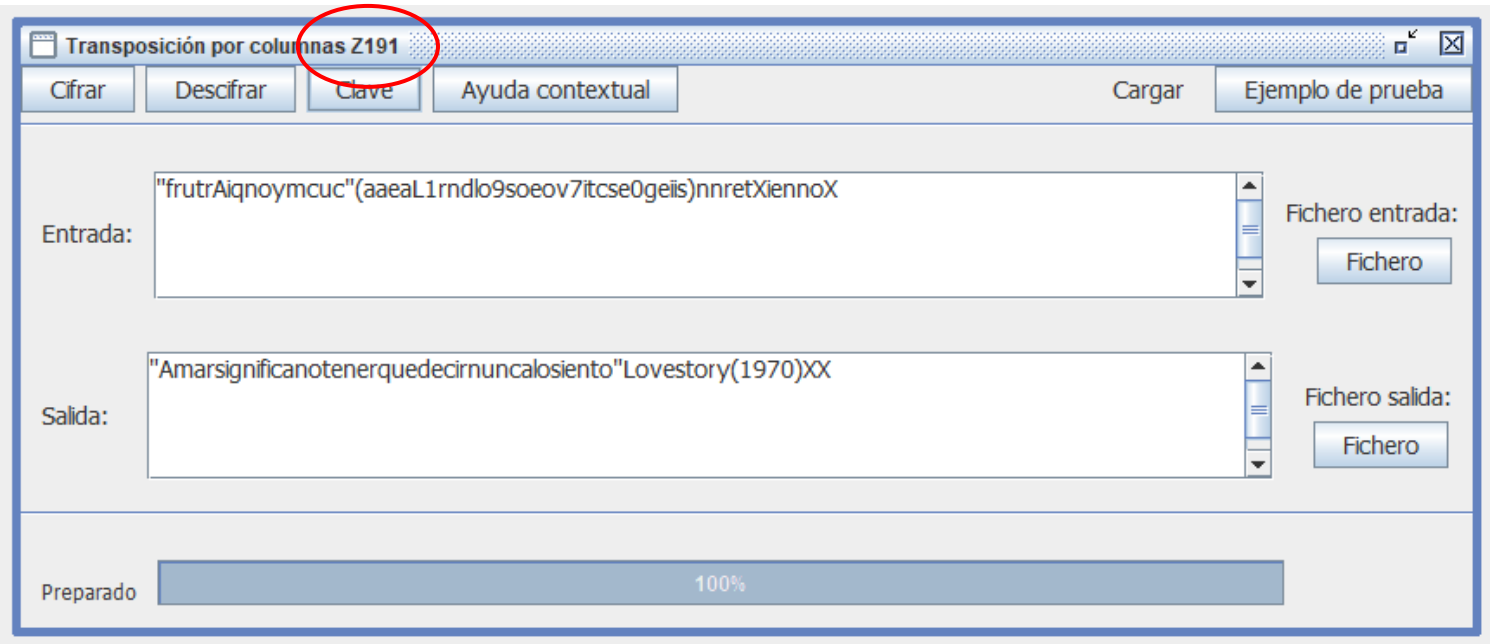
- C = EIETN OOOON MSCEU XASEX SVRX
- L = 24, NC = 6, NF = 24/6 = 4

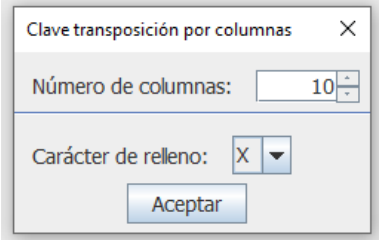
E	N	O	C	A	S
I	O	N	E	S	V
E	O	M	U	E	R
T	O	S	X	X	X

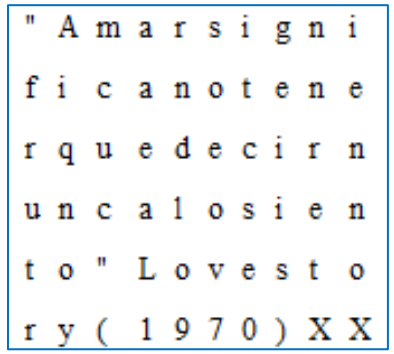
- M = EN OCASIONES VEO MUERTOS

Práctica de columnas con Criptoclásicos

- C = "frutrAiqn oymcuc"(aa eaL1rndlo9 soeov7itcs e0geiis)nn retXiennoX



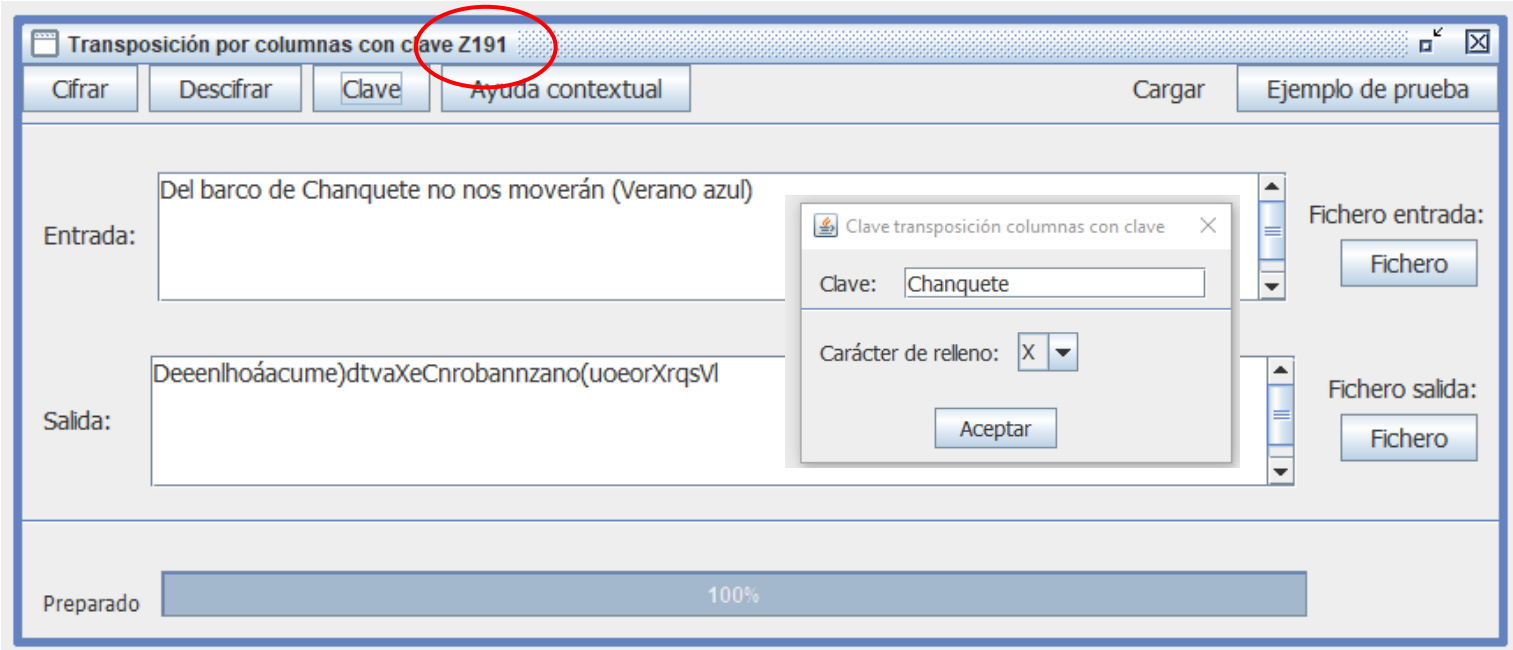




- M = “Amar significa no tener que decir nunca lo siento” Love story (1970)

Permutación de filas y columnas con clave

- M = Del barco de Chanquete no nos moverán (Verano azul)

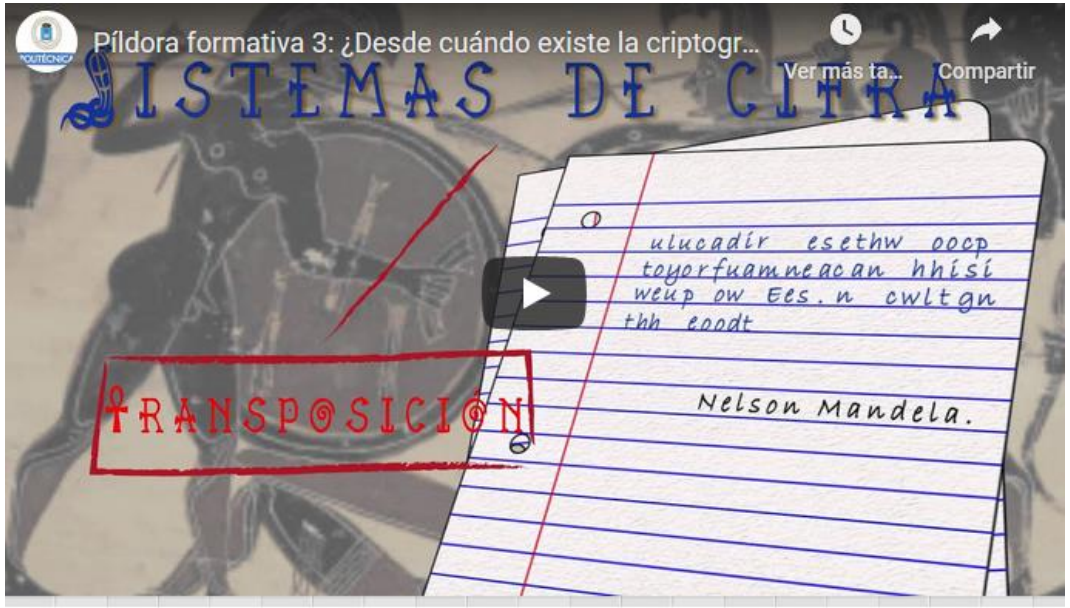


C	h	a	n	q	u	e	t	e
D	e	l	b	a	r	c	o	d
e	C	h	a	n	q	u	e	t
e	n	o	n	o	s	m	o	v
e	r	á	n	(	V	e	r	a
n	o	a	z	u	l	)	X	X

C	a	e	e	h	n	q	t	u
D	l	c	d	e	b	a	o	r
e	h	u	t	C	a	n	e	q
e	o	m	v	n	n	o	o	s
e	á	e	a	r	n	(	r	V
n	a	)	X	o	z	u	X	l

- C = Deeenlhoáa cume)dtvaX eCnrobannz ano(uoeorX rqsVl

Más información en píldoras Thoth



Píldora 3: ¿Desde cuándo existe la criptografía?

https://www.youtube.com/watch?v=jmvO3VoJeKg

Píldora 15: ¿Qué es la cifra por transposición o permutación?

https://www.youtube.com/watch?v=huIiPnr61sM

Conclusiones de la Lección 6.1a

- El cifrado mediante técnicas de permutación permite lograr el método de difusión, difundir las características del lenguaje
- Las letras del criptograma son las mismas que las del texto en claro, solo que están puestas en posiciones diferentes, están permutadas o transpuestas
- Existen diferentes técnicas de permutación. En esta lección hemos visto:
  - Cifrado por escítala
  - Cifrado por permutación de filas o columnas
  - Cifrado por permutación de filas o columnas con clave

Lectura recomendada

- Guion píldora formativa Thoth nº 3, ¿Desde cuándo existe la criptografía?, Jorge Ramió, 2014
  - https://www.criptored.es/thoth/material/texto/pildora003.pdf
- Guion píldora formativa Thoth nº 15, ¿Qué es la cifra por transposición o permutación?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora015.pdf
- Introducción a la seguridad informática y criptografía clásica, Lección 7 Algoritmos de cifra por transposición o permutación, MOOC Crypt4you, Jorge Ramió, 2016
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion7.html
- Cifrado por transposición, Wikipedia
  - https://es.wikipedia.org/wiki/Cifrado_por_transposici%C3%B3n

Class4crypt c4c6.1b

Módulo 6. Algoritmos de criptografía clásica

Lección 6.1b. Cifrado por permutación o transposición parte 2

6.1b.1. Cifrado tipo rail fence

6.1b.2. Cifrado por rejilla de Cardano

6.1b.3. Cifradores por permutación de bloques de texto

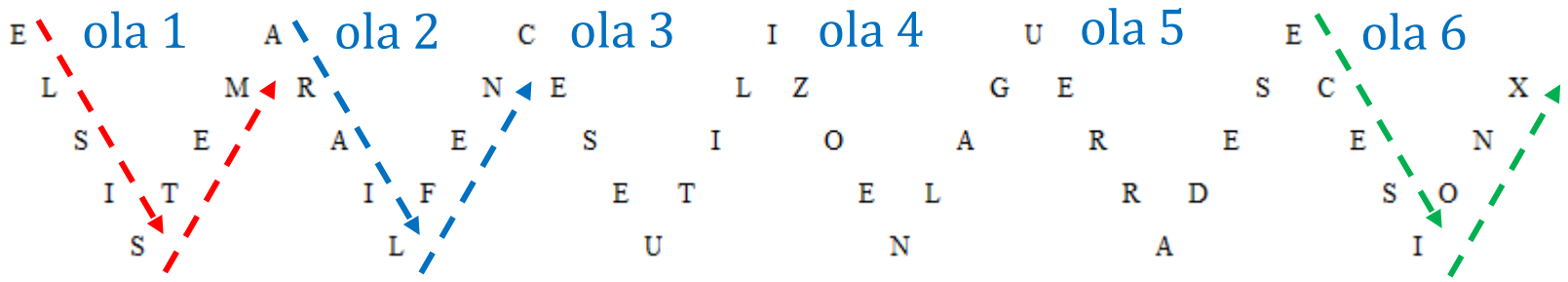
Class4crypt c4c6.1b Cifrado por permutación o transposición parte 2
https://www.youtube.com/watch?v=6vg059l2dR0

Permutaciones para lograr la difusión

- Recordando… clase anterior
- Esta técnica de cifra consiste en que las letras del texto en claro se redistribuyen dentro del criptograma sin sustituirlas, siguiendo unas reglas o patrón definidos
- Con esto se logra la difusión, es decir, difundir las características del texto en claro (del lenguaje) en todo el texto cifrado
- Ocultamos así la relación entre el texto en claro y el criptograma
- En esta segunda lección de cifradores por permutación, veremos sistemas de cifra con mayor seguridad que los de filas y columnas

Cifrando por rail fence

- La cifra consiste en escribir el texto en claro en varios renglones o raíles R, que van formando una figura tipo zigzag
- El criptograma es el resultado de la lectura de esos renglones
- Sea M = EL SISTEMA RAIL FENCE SE UTILIZO EN LA GUERRA DE SECESION (y se usará como clave R = 5 raíles)



Tamaño ola T = 2R – 2
T = 2*5-2 = 8
Letras texto claro L_p = 47
Número olas N = $\lceil L_p/T \rceil$
N = $\lceil 47/8 \rceil$ = 6

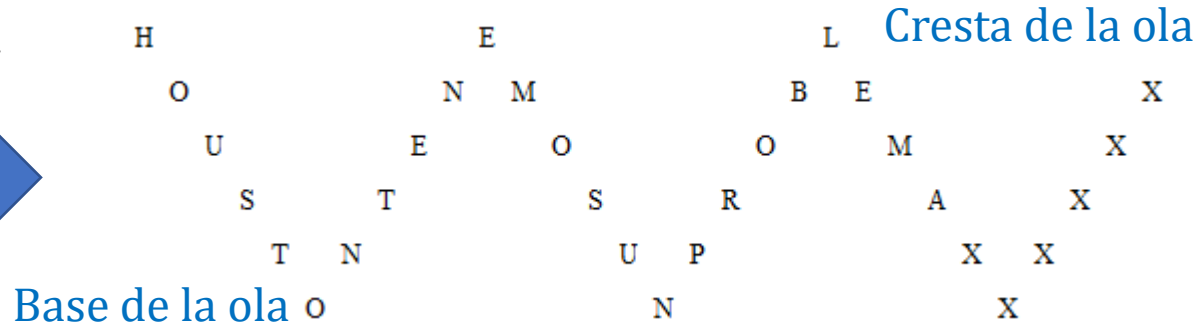
- C = EACIUE LMRNELZGESCX SEAESIOAREEN ITIFETELRDSO SLUNAI

Descifrando por rail fence

- Como conocemos la cantidad de letras L_c del criptograma y el número de raíles R, sabemos que habrá N olas con T = 2R-2 letras cada una, siendo N = L_c/T. Pero, ¿pero cómo formamos las olas?
- Sea C = HELON MBEXU EOOMX STSRA XTNUP XXONX
- Si la clave R = 6, entonces las olas serán de T = 2*6 - 2 = 10 letras
- Como L = 30 letras, entonces habrá N = L_c/T = 30/10 = 3 olas
- Entonces las 3 primeras letras corresponden a la cresta de la ola y las 3 últimas letras corresponde a la base de la ola
- Las 24 letras restantes (30 – 2*3) las dividimos entre el cuerpo de la ola, igual a 4: los 6 raíles menos el de la cresta y el de la base

Descifrando rail fence con Criptoclásicos

- Sabiendo que C = HEL ONMBEX UEOOMX STSRAX TNUPXX ONX y con los datos de la página anterior, hacemos lo siguiente
- Tomamos las 3 primeras letras de la cresta de la ola: HEL
- Tomamos los 3 últimas letras de la base de la ola: ONX
- Las restantes 24 letras ONMBEX UEOOMX STSRAX TNUPXX las dividimos entre el restante cuerpo de la olas: 6 – 2 = 4
- 24/4 = 6 letras por fila de ola
- Cuerpo 0 = ONMBEX
- Cuerpo 1 = UEOOMX
- Cuerpo 2 = STSRAX
- Cuerpo 3 = TNUPXX



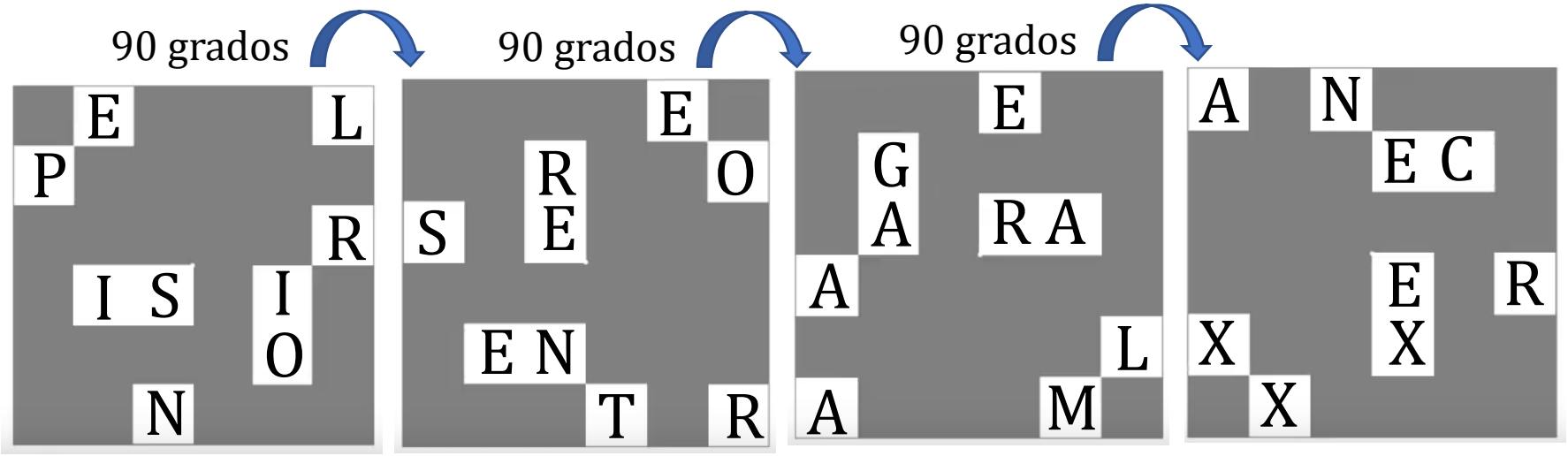
M = HOUSTON TENEMOS UN PROBLEMA

Rejilla de Cardano, ¿cifra o esteganografía?

- Sistema que propone Gerolamo Cardano en 1550
- Si se usa una rejilla mediante la cual se van poniendo las letras del texto en claro en las x celdas libres de la misma, luego ésta se gira 90 grados para seguir poniendo las otras letras del texto en claro en dichas celdas, y así se va repitiendo a lo largo de todo el texto, sin que se introduzca ningún otro texto que no sea el del secreto a cifrar, diremos que se trata de una cifra por permutación
- No obstante, si usamos las celdas de esa rejilla solamente para marcar las letras del secreto a ocultar dentro de otro texto del que no haya ninguna sospecha, entonces hablamos de esteganografía

Cifra con rejilla de Cardano 6x6 y 9 celdas

- M = EL PRISIONERO SE ENTREGARA AL AMANECER
- 9 espacios: ELPRISION EROSEENTR EGARAALAM ANECERXXX



A	E	N	E	E	L
P	G	R	E	C	O
S	A	E	R	A	R
A	I	S	E	I	R
X	E	N	X	O	L
A	X	N	T	M	R

- C = AENEEL PGRECO SAERAR AISEIR XENXOL AXNTMR

Permutación por reordenación de números

- La cifra anterior con rejilla, propia de aquellos años 1550, puede hacerse mediante la permutación de una lista de números
- En una matriz de nxm celdas o números, estos números se reordenan siguiendo un proceso conocido o de forma aleatoria
- En el ejemplo anterior, los 6x6 = 36 números de la rejilla se han reordenado de esta manera:
  - Estado 1, celdas: 2, 6, 7, 18, 20, 21, 23, 29, 33
  - Estado 2, celdas: 5, 9, 12, 13, 15, 26, 27, 34, 36
  - Estado 3, celdas: 4, 8, 14, 16, 17, 19, 30, 31, 35
  - Estado 4, celdas: 1, 3, 10, 11, 22, 24, 25, 28, 32

Permutación por bloques de texto y matriz

N	O	H	A	Y
M	A	L	Q	U
E	P	O	R	B
I	E	N	N	O
V	E	N	G	A

- M = NO HAY MAL QUE POR BIEN NO VENGA
- Desde el centro y formando una espiral según el sentido de las agujas de un reloj
- C = OLQRN NEPAO HAYUB OAGNE VIEMN
- Espiral sentido contrario a las agujas del reloj
- C = OLAPE NNRQA HONME IVENG AOBUY

- Por ejemplo, leyendo líneas en diagonal
- C = NMOEA HIPLA VEOQY ENRUN NBGOA
- C = AGONN BENRU VEOQY IPLAE AHMON

Sentido de la lectura

De izquierda a derecha

De derecha a izquierda

Complicando la permutación

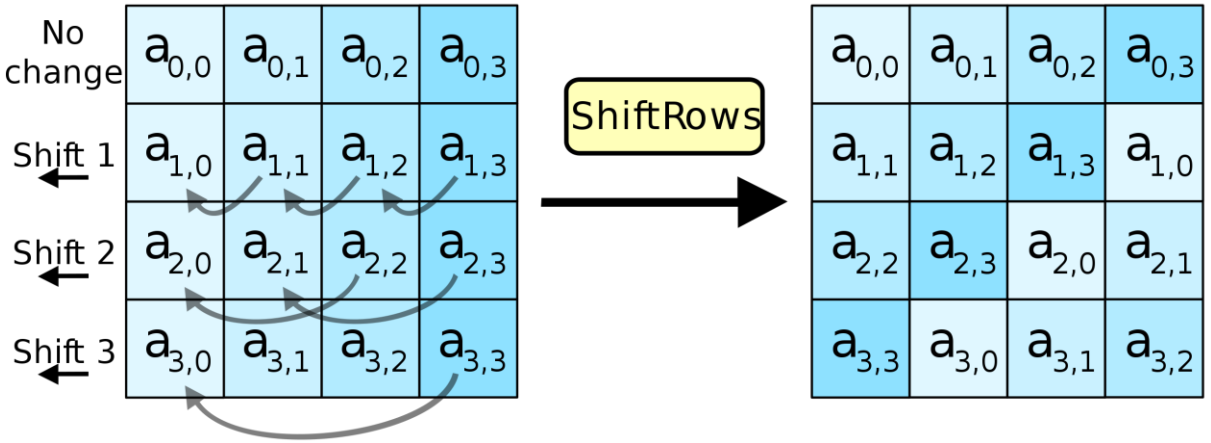
- Las figuras que pueden hacerse al escribir el texto en claro en una matriz son “infinitas”
- En algunos casos, esto consigue que la fortaleza de la cifra sea resistente durante muchos años
- Como el criptograma del “asesino del zodiaco”

Transposition scheme *Source: David Oranchak/YouTube*

Aunque sea criptografía por oscurantismo, hoy inaceptable, aquí el secreto ha tenido una fortaleza mayor que 50 años…

Permutación en la cifra moderna

IP							
58	50	42	34	26	18	10	2
60	52	44	36	28	20	12	4
62	54	46	38	30	22	14	6
64	56	48	40	32	24	16	8
57	49	41	33	25	17	9	1
59	51	43	35	27	19	11	3
61	53	45	37	29	21	13	5
63	55	47	39	31	23	15	7



- Algoritmo DES (Permutation tables)
- Algoritmo AES (ShiftRows)

Más información en píldoras Thoth

Píldora 15: ¿Qué es la cifra por transposición o permutación?

https://www.youtube.com/watch?v=huIiPnr61sM

Conclusiones de la Lección 6.1b

- El cifrado mediante técnicas de permutación permite lograr el método de difusión, difundir las características del lenguaje
- Las letras del criptograma son las mismas que las del texto en claro, solo que están puestas en posiciones diferentes, están permutadas o transpuestas
- Existen diferentes técnicas de permutación. En esta lección hemos visto:
  - Cifrado por permutación rail fence
  - Cifrado por permutación de rejilla de Cardano
  - Cifrado por permutación de secuencia de números en matrices
  - Cifrado por permutación usando figuras geométricas
- Los sistemas modernos como DES y AES usan permutaciones de bits y bytes

Lectura recomendada

- Guion píldora formativa Thoth nº 15, ¿Qué es la cifra por transposición o permutación?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora015.pdf
- Introducción a la seguridad informática y criptografía clásica, Lección 7 Algoritmos de cifra por transposición o permutación, MOOC Crypt4you, Jorge Ramió, 2016
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion7.html
- Transposition cipher, Wikipedia
  - https://en.wikipedia.org/wiki/Transposition_cipher
- Cifrado 340. Al asesino del Zodiaco se le van acabando los secretos, La Vanguardia, 2020
  - https://www.lavanguardia.com/cultura/20201215/6122813/asesino-zodiaco-secretos-cifrado-340-criptograma-descifrado-codigo.html

Class4crypt c4c6.2

Módulo 6. Algoritmos de criptografía clásica

Lección 6.2. Criptoanálisis a la cifra por permutación

6.2.1. Estadísticas del lenguaje: digramas o bigramas

6.2.2. Criptoanálisis por anagramación en cifras de filas y columnas

6.2.3. Dificultad del criptoanálisis por la complejidad del modo de cifra

Class4crypt c4c6.2 Criptoanálisis a la cifra por permutación
https://www.youtube.com/watch?v=Wu9dGHjj2xA

Estadísticas de digramas en el lenguaje

Frecuencias de digramas

Tabla de frecuencia de digramas en Español

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
A	0.291%	0.182%	1.014%	1.232%	0.519%	0.115%	0.078%	0.057%	0.152%	0.106%	0.002%	1.29%	0.607%	1.004%	0.02%	0.171%	0.809%	0.163%	1.418%	1.638%	0.322%	0.219%	0.162%	0.0%	0.03%	0.169%	0.015%
B	0.143%	0.001%	0.002%	0.004%	0.137%	0.003%	0.001%	0.0%	0.153%	0.004%	0.0%	0.141%	0.004%	0.001%	0.0%	0.192%	0.002%	0.0%	0.14%	0.005%	0.042%	0.022%	0.0%	0.001%	0.0%	0.002%	0.0%
C	0.783%	0.001%	0.184%	0.01%	0.328%	0.002%	0.001%	0.284%	1.362%	0.0%	0.001%	0.117%	0.001%	0.002%	0.0%	1.056%	0.005%	0.0%	0.091%	0.004%	0.325%	0.353%	0.0%	0.0%	0.0%	0.003%	0.0%
D	0.799%	0.006%	0.014%	0.046%	2.479%	0.002%	0.0%	0.001%	0.54%	0.003%	0.001%	0.009%	0.013%	0.006%	0.0%	0.934%	0.01%	0.025%	0.043%	0.017%	0.008%	0.218%	0.005%	0.003%	0.0%	0.003%	0.0%
E	0.546%	0.174%	0.928%	0.698%	0.459%	0.178%	0.234%	0.11%	0.195%	0.098%	0.002%	1.689%	0.447%	2.208%	0.144%	0.133%	0.422%	0.123%	1.437%	2.588%	0.316%	0.165%	0.119%	0.004%	0.118%	0.032%	0.071%
F	0.085%	0.001%	0.005%	0.003%	0.096%	0.005%	0.005%	0.0%	0.524%	0.0%	0.0%	0.024%	0.002%	0.001%	0.0%	0.109%	0.002%	0.0%	0.027%	0.004%	0.01%	0.047%	0.0%	0.0%	0.0%	0.001%	0.0%
G	0.194%	0.002%	0.003%	0.002%	0.131%	0.001%	0.001%	0.002%	0.058%	0.0%	0.0%	0.003%	0.001%	0.033%	0.0%	0.07%	0.002%	0.001%	0.3%	0.004%	0.0%	0.282%	0.001%	0.0%	0.001%	0.002%	0.0%
H	0.286%	0.0%	0.001%	0.0%	0.187%	0.0%	0.0%	0.0%	0.03%	0.0%	0.0%	0.001%	0.0%	0.001%	0.0%	0.065%	0.004%	0.0%	0.0%	0.0%	0.001%	0.005%	0.0%	0.0%	0.0%	0.001%	0.0%
I	0.291%	0.107%	0.718%	0.512%	0.557%	0.121%	0.249%	0.005%	0.009%	0.009%	0.0%	0.24%	0.442%	0.792%	0.0%	1.301%	0.133%	0.032%	0.29%	0.427%	0.277%	0.012%	0.167%	0.0%	0.002%	0.0%	0.162%
J	0.054%	0.0%	0.002%	0.001%	0.118%	0.0%	0.0%	0.0%	0.0%	0.0%	0.001%	0.001%	0.0%	0.0%	0.0%	0.043%	0.001%	0.0%	0.0%	0.0%	0.0%	0.049%	0.0%	0.0%	0.0%	0.0%	0.0%
K	0.002%	0.003%	0.002%	0.0%	0.006%	0.0%	0.001%	0.0%	0.0%	0.001%	0.0%	0.001%	0.001%	0.001%	0.0%	0.0%	0.001%	0.001%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
L	1.495%	0.05%	0.165%	0.118%	0.54%	0.091%	0.077%	0.008%	0.524%	0.004%	0.0%	0.18%	0.137%	0.117%	0.0%	0.793%	0.282%	0.057%	0.047%	0.172%	0.24%	0.111%	0.046%	0.0%	0.0%	0.012%	0.0%
M	0.281%	0.15%	0.003%	0.004%	0.588%	0.0%	0.001%	0.0%	0.476%	0.0%	0.001%	0.001%	0.002%	0.026%	0.0%	0.363%	0.385%	0.001%	0.0%	0.02%	0.0%	0.075%	0.002%	0.0%	0.001%	0.002%	0.0%
N	0.862%	0.034%	0.41%	0.667%	0.87%	0.191%	0.103%	0.012%	0.306%	0.011%	0.002%	0.282%	0.082%	0.048%	0.0%	0.473%	0.19%	0.049%	0.036%	0.279%	1.379%	0.303%	0.046%	0.0%	0.001%	0.03%	0.019%
Ñ	0.153%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.011%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
O	0.187%	0.178%	0.422%	0.955%	0.368%	0.066%	0.234%	0.045%	0.053%	0.014%	0.001%	0.44%	0.464%	1.685%	0.0%	0.044%	0.402%	0.098%	1.079%	2.06%	0.214%	0.056%	0.042%	0.003%	0.003%	0.069%	0.003%
P	0.535%	0.0%	0.141%	0.002%	0.486%	0.0%	0.011%	0.0%	0.077%	0.0%	0.0%	0.117%	0.001%	0.001%	0.0%	0.77%	0.006%	0.001%	0.767%	0.004%	0.034%	0.482%	0.0%	0.0%	0.0%	0.004%	0.0%
Q	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.001%	0.0%	0.0%	0.873%	0.0%	0.0%	0.0%	0.0%	0.0%
R	1.495%	0.013%	0.181%	0.179%	1.469%	0.029%	0.022%	0.008%	0.541%	0.006%	0.001%	0.261%	0.299%	0.047%	0.0%	1.11%	0.086%	0.04%	0.17%	0.174%	0.326%	0.178%	0.03%	0.0%	0.001%	0.024%	0.003%
S	0.58%	0.066%	0.419%	0.634%	1.53%	0.108%	0.037%	0.045%	0.699%	0.011%	0.007%	0.182%	0.158%	0.076%	0.0%	0.655%	0.523%	0.22%	0.067%	0.324%	1.108%	0.337%	0.054%	0.0%	0.001%	0.132%	0.0%
T	1.332%	0.002%	0.014%	0.011%	1.171%	0.008%	0.0%	0.001%	0.711%	0.001%	0.001%	0.002%	0.002%	0.001%	0.0%	0.743%	0.005%	0.002%	0.574%	0.012%	0.006%	0.174%	0.001%	0.008%	0.012%	0.003%	0.0%
U	0.318%	0.02%	0.215%	0.057%	1.26%	0.014%	0.027%	0.0%	0.258%	0.004%	0.0%	0.254%	0.243%	0.621%	0.0%	0.037%	0.095%	0.001%	0.185%	0.186%	0.14%	0.001%	0.007%	0.0%	0.0%	0.031%	0.015%
V	0.248%	0.0%	0.0%	0.0%	0.255%	0.0%	0.001%	0.0%	0.107%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.071%	0.004%	0.0%	0.0%	0.001%	0.0%	0.003%	0.0%	0.001%	0.0%	0.0%	0.0%
W	0.014%	0.0%	0.001%	0.0%	0.0%	0.0%	0.0%	0.0%	0.001%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%	0.001%	0.0%	0.0%	0.0%	0.0%	0.001%	0.0%	0.0%
X	0.013%	0.0%	0.01%	0.0%	0.004%	0.0%	0.0%	0.0%	0.068%	0.0%	0.001%	0.001%	0.0%	0.001%	0.0%	0.001%	0.029%	0.0%	0.003%	0.002%	0.034%	0.0%	0.0%	0.0%	0.011%	0.002%	0.0%
Y	0.072%	0.009%	0.03%	0.046%	0.071%	0.017%	0.007%	0.005%	0.002%	0.0%	0.0%	0.029%	0.014%	0.011%	0.0%	0.024%	0.044%	0.036%	0.013%	0.045%	0.014%	0.023%	0.01%	0.0%	0.0%	0.001%	0.001%
Z	0.187%	0.0%	0.03%	0.003%	0.007%	0.0%	0.001%	0.0%	0.009%	0.0%	0.0%	0.003%	0.001%	0.0%	0.0%	0.016%	0.001%	0.022%	0.0%	0.006%	0.001%	0.001%	0.0%	0.0%	0.0%	0.001%	0.0%

Digrama: ES Frecuencia: 2.588%

- Otros digramas frecuentes en castellano: DE (2,48%), EN (2,21%), OS (2,06%), EL (1,69%), ON (1,69%), AS (1,64%), SE (1,53%)

Anagramación en cifra de filas o columnas

- La técnica de anagramación consiste en la comparación de bloques de dos letras del criptograma con el objeto de encontrar digramas comunes del lenguaje, que se han destruidos por la transposición
- Se elige una cadena de letras sucesivas en el inicio del criptograma llamada ventana, que contenga al menos 4 letras para poder así aplicar estadísticas con cierta confianza
- La ventana se va desplazando por el criptograma en incrementos de una letra o una posición hacia la derecha, y se ven los digramas que se van formando entre esa ventana y el texto cifrado
- Una media alta y una desviación típica baja indican posible acierto

Comparando digramas por anagramación

- C = LPUOO DOADR ENINC AOTTS ECSEF NDCAI CRARI FOSRA TOAGO IPCAC ILEOC ETSRN FEINI LUPMI UEDAX RROSO AMUPL SCEMX AMNPN SNEEM ANAAX (cifrado por filas)
- Sea la ventana de 8 letras: LPUOODOA

	L	P	U	O	O	D	O	A
+ 1 espacio →	P	U	O	O	D	O	A	D

- Diagramas LP, PU, UO, OO, OD, DO, OA y AD
- Se mira en la tabla la frecuencia de cada uno de los digramas, se calcula la media de esas frecuencias y la desviación típica

Comparando digramas (cont.)

	L	P	U	O	O	D	O	A
+ 2 espacios →	U	O	O	D	O	A	D	R

	L	P	U	O	O	D	O	A
+ 3 espacios →	O	O	D	O	A	D	R	E

	L	P	U	O	O	D	O	A
+ 4 espacios →	O	D	O	A	D	R	E	N

- ¿Es necesario hacer todos los desplazamientos posibles?
- No. Se tomarán en cuenta solamente periodos en los que se pueda dividir la longitud del criptograma. Puesto que el criptograma del ejemplo tiene 105 letras, estos números son: 3, 5, 7, 15, 21 y 35

Medias y desviaciones en 3, 5, 7, 15, 21, 35

- Paso 3: ventana LPUOODOA con OODOADRE (digramas LO, PO, UD, OO, OA, DD, OR, AE)
  - media 0,43688 y desviación típica 0,38255
- Paso 5: ventana LPUOODOA con DOADRENI (digramas LD, PO, UA, OD, OR, DE, ON, AI)
  - media 0,94450 y desviación típica 0,76284
- Paso 7: ventana LPUOODOA con ADRENINC (digramas LA, PD, UR, OE, ON, DI, ON, AC)
  - media 0,87175 y desviación típica 0,64453
- Paso 15: ventana LPUOODOA con AOTTSECS (digramas LA, PO, UT, OT, OS, DE, OC, AS)
  - media 1,15225 y desviación típica 0,83082
- Paso 21: ventana LPUOODOA con CSEFNDCA (digramas LC, PS, UE, OF, ON, DD, OC, AA)
  - media 0,49238 y desviación típica 0,58990
- Paso 35: ventana LPUOODOA con FOSRATOA (digramas LF, PO, US, OR, OA, DT, OO, AA)
  - media 0,33200 y desviación típica 0,36053

Solución del criptoanálisis

- Paso 15: media 1,15225 y desviación típica 0,83082 (el mejor valor encontrado)
- Como el criptograma tiene 105 letras y el mejor resultado se obtiene para el periodo o paso 15, entonces se estima que el número de filas puede ser 105/15 = 7
- Se descifra el criptograma cifrado por filas con clave igual 7 (y 15 columnas)

L	P	U	O	O	D	O	A	D	R	E	N	I	N	C
A	O	T	T	S	E	C	S	E	F	N	D	C	A	I
C	R	A	R	I	F	O	S	R	A	T	O	A	G	O
I	P	C	A	C	I	L	E	O	C	E	T	S	R	N
F	E	I	N	I	L	U	P	M	I	U	E	D	A	X
R	R	O	S	O	A	M	U	P	L	S	C	E	M	X
A	M	N	P	N	S	N	E	E	M	A	N	A	A	X

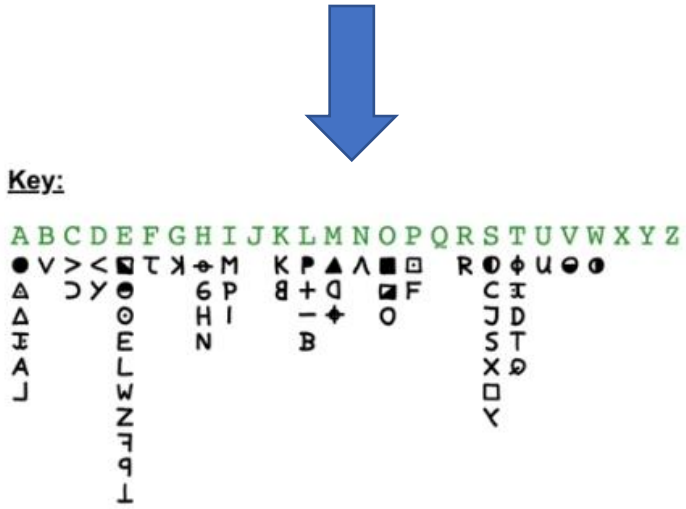
- M = LA CIFRA POR PERMUTACION O TRANSPOSICION DE FILAS O COLUMNAS SE PUEDE ROMPER FACILMENTE USANDO TECNICAS DE ANAGRAMACION

¿Y si la cifra de filas o columnas usa clave?

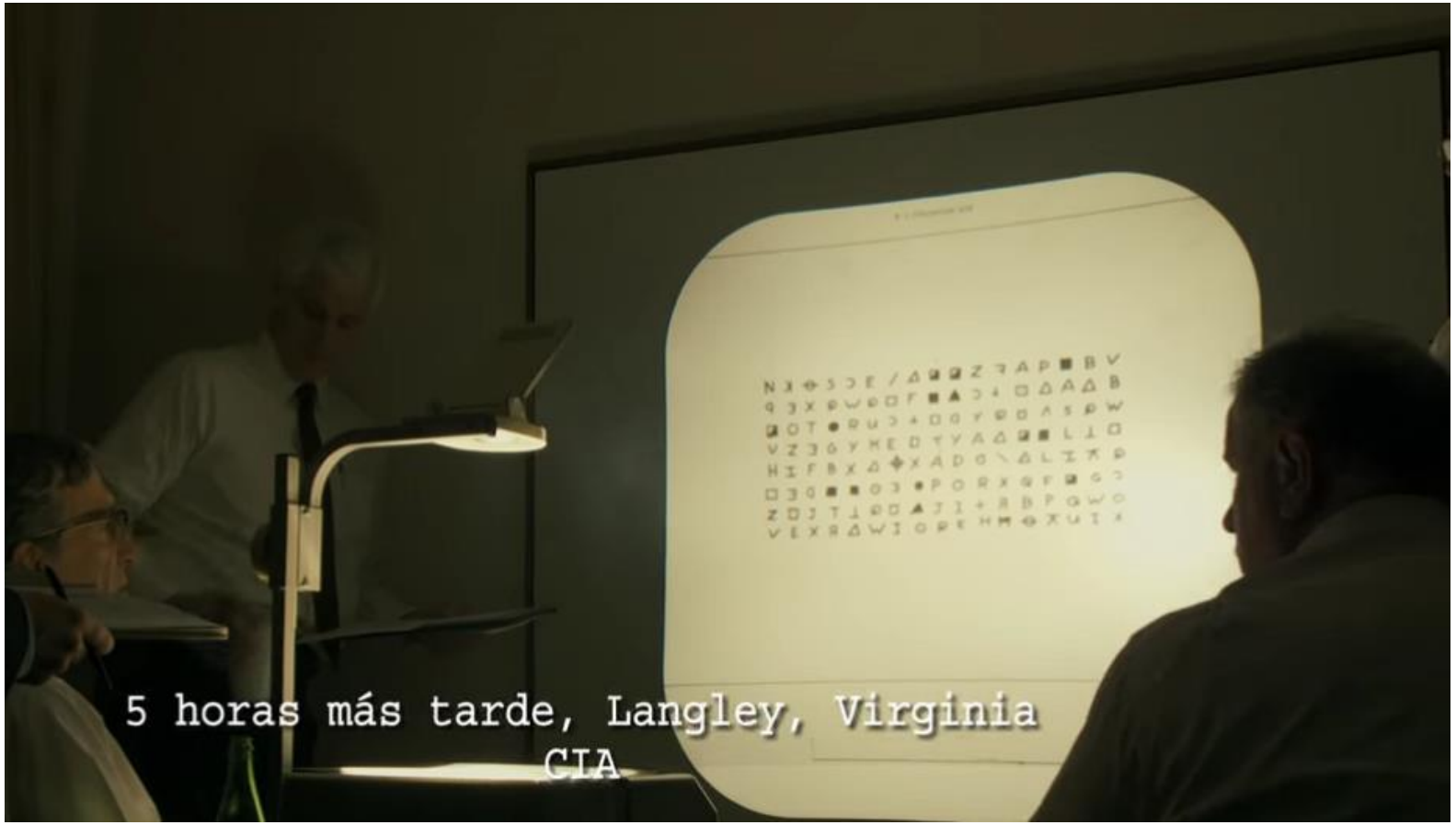
- Si la cifra por permutación de filas o columnas usa una clave, el criptoanálisis por anagramación se complica puesto que ya no puede usarse esa ventana que se va desplazando por todo el criptograma buscando digramas frecuentes
- También habrá que buscar digramas frecuentes en el criptograma, pero éstos ya no estarán de una forma tan ordenada como en el ejemplo anterior sino desperdigados, lo que complica el ataque
- Lo mismo sucede si la transposición se realiza por otros métodos como bloques de texto, la rejilla de Cardano o rail fence. En este último caso, resulta más simple hacer un ataque por fuerza bruta porque normalmente el número de raíles suele ser pequeño

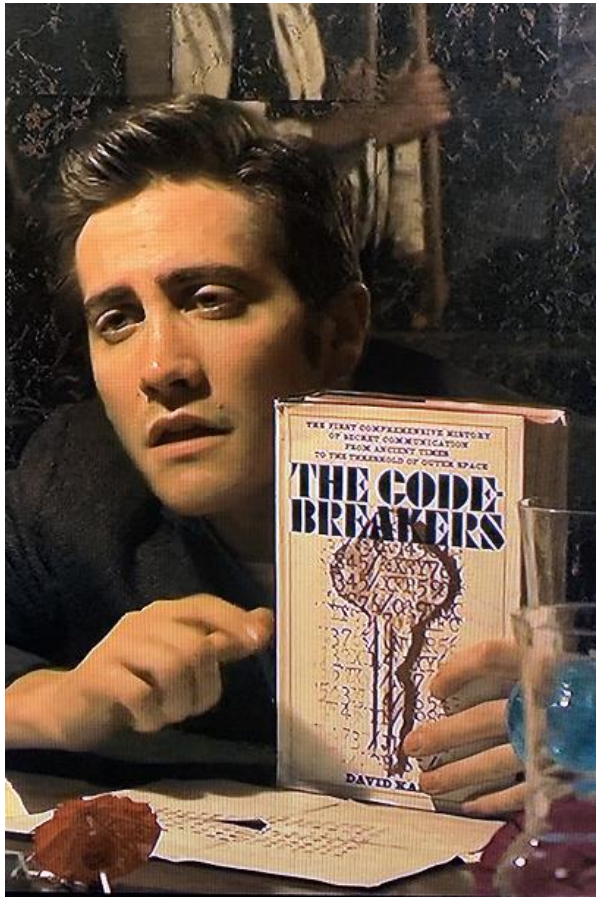
A veces un criptoanálisis no es tan simple

- El criptograma del zodiaco, con 340 caracteres en 17 columnas y 20 filas, se criptoanaliza en diciembre de 2020, 51 años después…
- Con un alfabeto de 63 caracteres, usa sustitución por homófonos y después transposición



Y una película: Zodiac (2007)





Más información en píldoras Thoth

Píldora 15: ¿Qué es la cifra por transposición o permutación?

https://www.youtube.com/watch?v=huIiPnr61sM

Conclusiones de la Lección 6.2

- La técnica para criptoanalizar sistemas de cifra clásica por transposición o permutación se conoce como anagramación
- La anagramación consiste en buscar digramas o bigramas frecuentes del texto en claro dentro del criptograma, que pudiesen ser parte del texto en claro
- Para la cifra por permutación de filas o columnas sin clave, podemos usar una ventana de unas 5 a 10 letras que vamos recorriendo por todo el criptograma buscando digramas que tengan una media alta y una baja desviación típica
- Esto se hace para que un diagrama demasiado frecuente como puede ser DE, EN, OS, etc. no fuerce a que la media sea muy alta y por ejemplo encontremos en ese ventana otro diagrama QQ cuya frecuencia en castellano es cero
- En otros sistemas de permutación, este ataque podría ser más complejo

Lectura recomendada (1/2)

- Guion píldora formativa Thoth nº 15, ¿Qué es la cifra por transposición o permutación?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora015.pdf
- Introducción a la seguridad informática y criptografía clásica, Lección 7 Algoritmos de cifra por transposición o permutación, MOOC Crypt4you, Jorge Ramió, 2016
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion7.html
- Elements of Cryptanalysis, Washington Government Printing Office, 1923 (approv NSA 2014)
  - https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/friedman-documents/publications/FOLDER_237/41760269079943.pdf
- Cifrado 340. Al asesino del Zodiaco se le van acabando los secretos, La Vanguardia, 15/12/2020
  - https://www.lavanguardia.com/cultura/20201215/6122813/asesino-zodiaco-secretos-cifrado-340-criptograma-descifrado-codigo.html

Lectura recomendada (2/2)

- Analysis of the Zodiac 340-cipher, Thang Đào, San Jose State University, 2008
  - http://scholarworks.sjsu.edu/cgi/viewcontent.cgi?article=4566&context=etd_theses
- Let's Crack Zodiac, David Oranchak, 2020
  - Episode 1 - Graysmith: https://www.youtube.com/watch?v=N_Oh4snhF70
  - Episode 2 - Bigrams: https://www.youtube.com/watch?v=uyH9-AHz-HY
  - Episode 3 - The 408: https://www.youtube.com/watch?v=57o8g3d6lSw
  - Episode 4 - National Cryptologic Museum: https://www.youtube.com/watch?v=ZY1Y5xyKcBw
  - Episode 5 - The 340 Is Solved!: https://www.youtube.com/watch?v=-1oQLPRE21o
  - Episode 6 - How was the 340 made?: https://www.youtube.com/watch?v=UfrFSrI05hk
- Is David Oranchak's Zodiac Killer Cipher Z-340 Solution Correct?, Nils Kopal, 2020
  - https://www.youtube.com/watch?v=hIrOftXgibg

Class4crypt c4c6.3

Módulo 6. Algoritmos de criptografía clásica

Lección 6.3. Cifrado por sustitución

6.3.1. Técnicas de sustitución para lograr la confusión

6.3.2. Sustitución monoalfabética

6.3.3. Cifrado por sustitución monoalfabética con desplazamiento o César

6.3.4. Cifrado por sustitución monoalfabética con decimación

6.3.5. Cifrado por sustitución monoalfabética afín

6.3.6. Introducción a la sustitución polialfabética

Class4crypt c4c6.3 Cifrado por sustitución
https://www.youtube.com/watch?v=uRp26lF0rkM

Sustituciones para lograr la confusión

- Esta técnica de cifra consiste en que las letras del texto en claro se sustituyen o cambian por otras letras, o símbolos, siguiendo unas reglas o patrón definidos
- Con esto se logra la confusión, es decir, que al criptoanalista no le sea fácil encontrar una relación entre el criptograma y la clave
- Si sustituimos cada letra del alfabeto en claro por otra única letra (o símbolo) del alfabeto de cifrado, estaremos realizando una cifra de tipo monoalfabética. Si, por el contrario, cada letra del alfabeto en claro se sustituye por diferentes letras (o símbolos) del alfabeto de cifrado, la cifra se conoce como polialfabética

Sustitución monoalfabética

- Lógicamente, el criptograma no tendrá los mismos elementos que el texto en claro y ahora sí habrá que hacer una operación modular

									10										20								
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	Codificación letras
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	Alfabeto en claro
G	W	T	Ñ	H	S	P	Y	B	X	I	Z	U	L	V	K	M	R	A	F	D	C	E	O	N	J	Q	Alfabeto cifrado 1
J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	Alfabeto cifrado 2

- El alfabeto 1 tiene las 27 letras distribuidas de forma aleatoria y el alfabeto 2 muestra un desplazamiento constante de 9 espacios (J)
- Menos común sería un alfabeto de cifrado con signos, por ejemplo:
  - A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
  -

Cifrado ROT 13

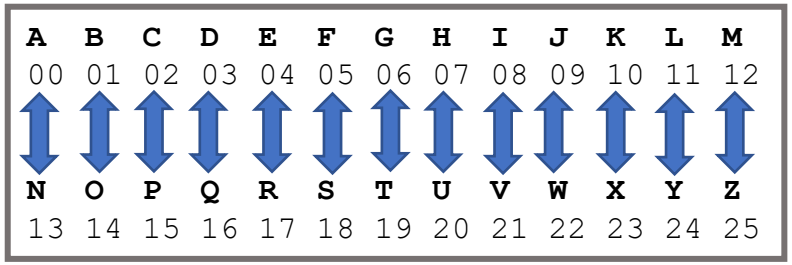
- Sistema muy simple usado en foros de noticias y apropiado para el alfabeto inglés de 26 letras

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Alfabeto inglés

Cifrado: $c_i = m_i + 13 \bmod 26$

Descifrado: $m_i = c_i + 13 \bmod 26$



- Cifrado:
  - HELLO MARY LOU = URYYB ZNELY BH
- Descifrado:
  - EVPXL ARYFB A = RICKY NELSON

Cifrado por desplazamiento mod 27

- Cifrado: $c_i = m_i + b \bmod 27$ $\quad$ $0 < b < 27$ porque siempre habrá inverso

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Desplazamiento b = 5																										
F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E

Alfabeto español

- Texto en claro: BUENOS DIAS PRINCESA
- Criptograma: GZJRT XINFX UWNRH JXF
  - B + 5 mod 27 = 1 + 5 mod 27 = 6 = G $\quad$ U + 5 mod 27 = 21 + 5 mod 27 = 26 = Z
  - E + 5 mod 27 = 4 + 5 mod 27 = 9 = J $\quad$ N + 5 mod 27 = 13 + 5 mod 27 = 18 = R
  - O + 5 mod 27 = 15 + 5 mod 27 = 20 = T $\quad$ S + 5 mod 27 = 19 + 5 mod 27 = 24 = X …

Descifrado por desplazamiento mod 27

- Descifrado: $m_i = c_i - b \bmod 27$ $\qquad$ $m_i = c_i + (27 - b) \bmod 27$

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**Ñ**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Desplazamiento b = 3 (para descifrar usamos - 3 o bien 27 - 3 = + 24)																										
D	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**Ñ**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**	**A**	**B**	**C**

Alfabeto español

- Criptograma: ÑDYLG DHVEH ÑÑD
- Texto en claro: LA VIDA ES BELLA
  - Ñ – 3 mod 27 = 14 – 3 mod 27 = 11 = L $\qquad$ D – 3 mod 27 = 3 – 3 mod 27 = 0 = A
  - Y – 3 mod 27 = 25 – 3 mod 27 = 22 = V $\qquad$ L – 3 mod 27 = 11 – 3 mod 27 = 8 = I
  - G – 3 mod 27 = 6 – 3 mod 27 = 3 = D $\qquad$ D – 3 mod 27 = 3 – 3 mod 27 = 0 = A …

Cifrado por decimación mod 27

- Cifrado: $c_i = a*m_i \bmod 27$ mcd (a, 27) = 1 parar asegurar el inverso

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Decimación a = 4 (valores de a = 2,4,5,7,8,10,11,13,14,16,17,19,20,22,23,25,26)																										
A	E	I	M	P	T	X	B	F	J	N	Q	U	Y	C	G	K	Ñ	R	V	Z	D	H	L	O	S	W

Alfabeto español

- Texto en claro: SAYONARA BABY
- Criptograma: VASGY ARAEA ES
  - 4 * S mod 27 = 4 * 19 = 76 mod 27 = 22 = V 4 * A mod 27 = 4 * 0 mod 27 = 0 = A
  - 4 * Y mod 27 = 4 * 25 = 100 mod 27 = 19 = S 4 * O mod 27 = 4 * 15 = 60 mod 27 = 6 = G
  - 4 * N mod 27 = 4 * 13 = 52 mod 27 = 25 = Y 4 * A mod 27 = 4 * 0 mod 27 = 0 = A …

Descifrado por decimación mod 27

- Ecuación de descifrado: $m_i = c_i * \text{inv}(a, 27) \bmod 27$

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**Ñ**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Decimación = 2 (para descifrar usamos inv (2, 27) = 14)																										
A	**C**	**E**	**G**	**I**	**K**	**M**	**Ñ**	**P**	**R**	**T**	**V**	**X**	**Z**	**B**	**D**	**F**	**H**	**J**	**L**	**N**	**O**	**Q**	**S**	**U**	**W**	**Y**

Alfabeto español

- Criptograma: NIJXP ZANDJ
- Texto en claro: TERMINATOR
  - N * 14 mod 27 = 13 * 14 = 182 mod 27 = 20 = T
  - I * 14 mod 27 = 8 * 14 = 112 mod 27 = 4 = E
  - J * 14 mod 27 = 9 * 14 = 126 mod 27 = 18 = R …

Cifrado afín mod 27

- Cifrado: $c_i = a*m_i + b \bmod 27$

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Decimación a = 7, desplazamiento b = 10																										
K	Q	X	E	L	R	Y	F	M	S	Z	G	N	T	A	H	Ñ	U	B	I	O	V	C	J	P	W	D

Alfabeto español

- Texto en claro: TOCALA OTRA VEZ SAM
- Criptograma: OHXKG KHOBK CLDIK N
  - 7 * T + 10 mod 27 = 7 * 20 + 10 = 150 mod 27 = 15 = O
  - 7 * O + 10 mod 27 = 7 * 15 + 10 = 115 mod 27 = 7 = H
  - 7 * C + 10 mod 27 = 7 * 2 + 10 = 24 mod 27 = 24 = X …

Descifrado afín mod 27

- Ecuación de descifrado: $m_i = (c_i - b) * \text{inv}(a, 27) \bmod 27$

A	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**Ñ**	**O**	**P**	**Q**	**R**	**S**	**T**	**U**	**V**	**W**	**X**	**Y**	**Z**
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
Decimación a = 4, desplazamiento b = 5 (para descifrar usamos inv (4, 27) = 7)																										
F	**J**	**N**	**Q**	**U**	**Y**	**C**	**G**	**K**	**Ñ**	**R**	**V**	**Z**	**D**	**H**	**L**	**O**	**S**	**W**	**A**	**E**	**I**	**M**	**P**	**T**	**X**	**B**

Alfabeto español

- Criptograma: NFAFJ VFDNF
- Texto en claro: CASABLANCA
  - (N – 5) * 7 mod 27 = (13 – 5) * 7 mod 27 = 56 mod 27 = 2 = C
  - (F – 5) * 7 mod 27 = (5 – 5) * 7 mod 27 = 0 mod 27 = 0 = A
  - (A – 5) * 7 mod 27 = (0 – 5) * 7 mod 27 = -35 mod 27 = 19 = S …

Sustitución polialfabética

- Cada letra del texto en claro se sustituye por dos o más elementos en el criptograma, según una clave o mediante homófonos
- Sustitución polialfabética con clave
- Si la clave fuese la palabra MITO, tenemos estos alfabetos de cifra:

									10											20							
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	Codificación letras
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	Alfabeto en claro
M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	Alfabeto cifrado M
I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	Alfabeto cifrado I
T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	Alfabeto cifrado T
O	P	Q	R	S	T	U	V	W	X	Y	Z	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	Alfabeto cifrado O

- Algoritmos: Vigenère, Beaufort, Clave continua, Gronsfeld

Sustitución por homófonos

- Homófono RAE: Dicho de una palabra que suena igual que otra, pero que tiene distinto significado y puede tener distinta grafía. Ejemplos: ola/hola, tubo/tuvo, río-verbo/río-sustantivo, etc.
- Podemos usar números o signos, asignando más homófonos a las letras más frecuentes del alfabeto para destruir las estadísticas:

Otros caracteres del teclado usados como homófonos

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z	Alfabeto en claro
0	2	4	6	8	1	3	5	7	9	º	ª	!	\	|	@	·	#	$	~	%	&	/	(	)	=	?	Alfabeto de cifrado
`				¿				¡							^				*								
[				]											+												
{				}											;												
+				ç																							
				:																							

Ejemplo de cifrado:
HOLA AMIGOS = 5@ª0` !73^~

Efecto de la sustitución polialfabética

- Con la sustitución monoalfabética NO se logra enmascarar las estadísticas y redundancia del lenguaje en el criptograma

- Con la sustitución polialfabética SÍ se logra enmascarar las estadísticas y redundancia del lenguaje en el criptograma

Sustitución en la cifra moderna

- Las cajas S (Substitution boxes) en DES, cambian 6 bits por 4 bits en cada una de las ocho cajas S (6 bytes los convierte a 4 bytes)

Bits 7 al 12: 101100 → Salida: 1101

S2	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	15	1	8	14	6	11	3	4	9	7	2	13	12	0	5	10
1	3	13	4	7	15	2	8	14	12	0	1	10	6	9	11	5
2	0	14	7	11	10	4	13	1	5	8	12	6	9	3	2	15
3	13	8	10	1	3	15	4	2	11	6	7	12	0	5	14	9

- La función SubBytes en AES cambia un byte por otro byte, SubBytes 7D = FF

	0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F
0	63	7C	77	7B	F2	6B	6F	C5	30	01	67	2B	FE	D7	AB	76
1	CA	82	C9	7D	FA	59	47	F0	AD	D4	A2	AF	9C	A4	72	C0
2	B7	FD	93	26	36	3F	F7	CC	34	A5	E5	F1	71	D8	31	15
3	04	C7	23	C3	18	96	05	9A	07	12	80	E2	EB	27	B2	75
4	09	83	2C	1A	1B	6E	5A	A0	52	3B	D6	B3	29	E3	2F	84
5	53	D1	00	ED	20	FC	B1	5B	6A	CB	BE	39	4A	4C	58	CF
6	D0	EF	AA	FB	43	4D	33	85	45	F9	02	7F	50	3C	9F	A8
7	51	A3	40	8F	92	9D	38	F5	BC	B6	DA	21	10	FF	F3	D2
8	CD	0C	13	EC	5F	97	44	17	C4	A7	7E	3D	64	5D	19	73
9	60	81	4F	DC	22	2A	90	88	46	EE	B8	14	DE	5E	0B	DB
A	E0	32	3A	0A	49	06	24	5C	C2	D3	AC	62	91	95	E4	79
B	E7	C8	37	6D	8D	D5	4E	A9	6C	56	F4	EA	65	7A	AE	08
C	BA	78	25	2E	1C	A6	B4	C6	E8	DD	74	1F	4B	BD	8B	8A
D	70	3E	B5	66	48	03	F6	0E	61	35	57	B9	86	C1	1D	9E
E	E1	F8	98	11	69	D9	8E	94	9B	1E	87	E9	CE	55	28	DF
F	8C	A1	89	0D	BF	E6	42	68	41	99	2D	0F	B0	54	BB	16

Más información en píldoras Thoth (1/2)

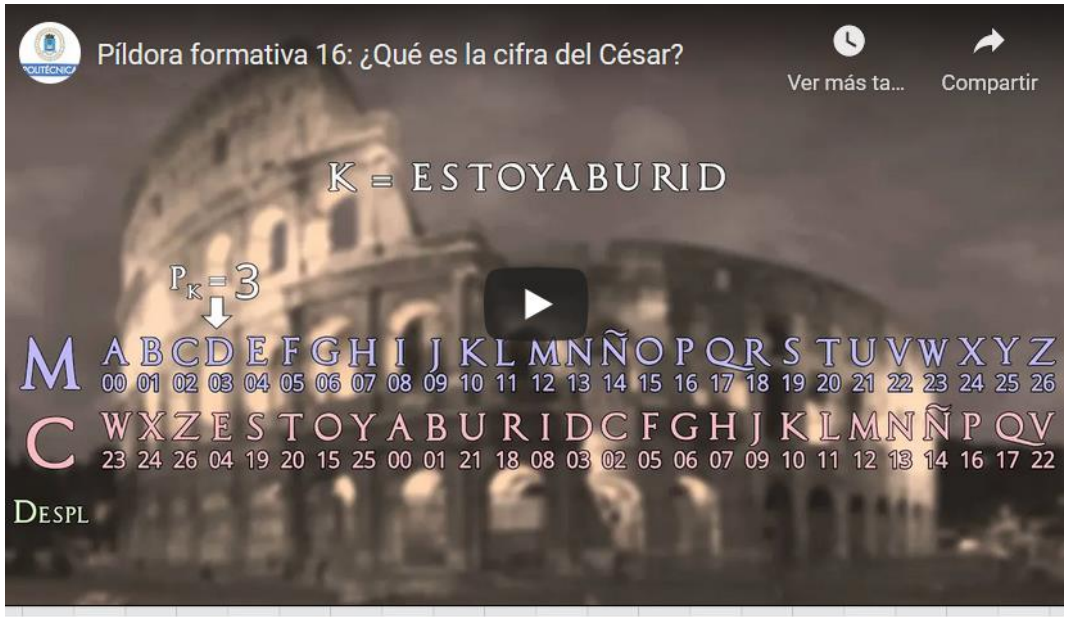
Píldora 13: ¿Qué es la cifra por sustitución monoalfabética?

https://www.youtube.com/watch?v=sSL04hd3TDU

Píldora 14: ¿Qué es la cifra por sustitución polialfabética?

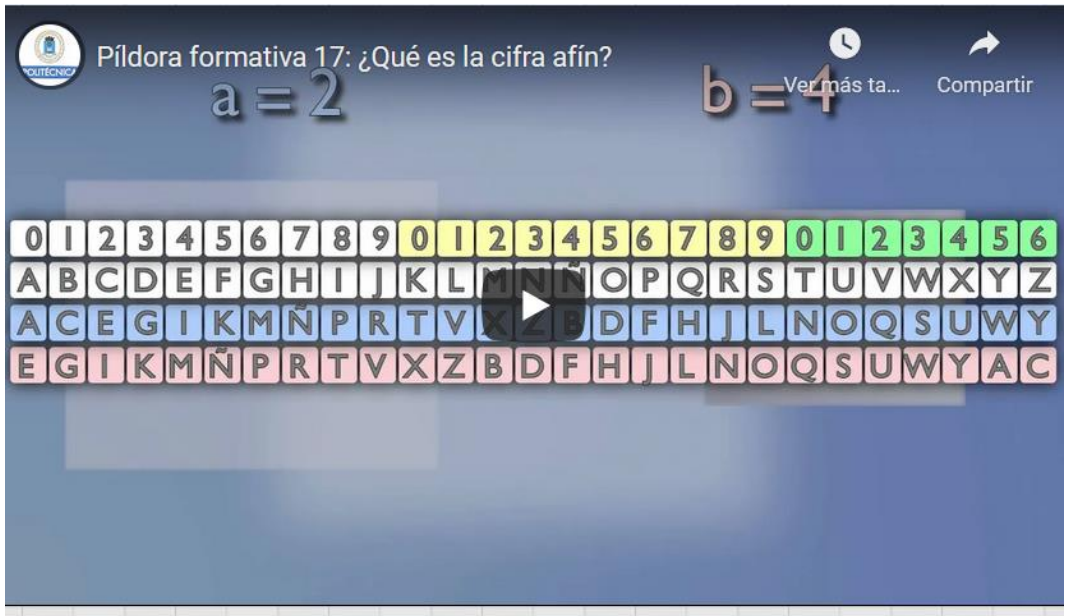
https://www.youtube.com/watch?v=VgmHd1ECkPA

Más información en píldoras Thoth (2/2)



Píldora 16: ¿Qué es la cifra del César?

https://www.youtube.com/watch?v=4vn6SIItjy0



Píldora 17: ¿Qué es la cifra afín?

https://www.youtube.com/watch?v=owmLwmlvnrU

Conclusiones de la Lección 6.3

- Para lograr confusión aplicamos operaciones de sustitución al texto en claro
- La sustitución puede ser monoalfabética (cada letra se sustituye por un único símbolo) o polialfabética (cada letra se sustituye por dos o más símbolos)
- La sustitución monoalfabética no enmascara las estadísticas del lenguaje en el criptograma, en cambio la polialfabética sí lo hace
- Entre los algoritmos más conocidos por sustitución monoalfabética se encuentran ROT 13, desplazamiento simple o César, decimación y afín
- Entre los algoritmos más conocidos por sustitución polialfabética se encuentran Vigenère, Beaufort, clave continua y Gronsfeld
- Otra forma de lograr una cifra polialfabética es mediante el uso de homófonos
- Sistemas de cifra moderna como AES y DES también usan sustitución binaria

Lectura recomendada (1/2)

- Guion píldora formativa Thoth nº 13, ¿Qué es la cifra por sustitución monoalfabética?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora013.pdf
- Guion píldora formativa Thoth nº 14, ¿Qué es la cifra por sustitución polialfabética?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora014.pdf
- Guion píldora formativa Thoth nº 16, ¿Qué es la cifra del César?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora016.pdf
- Guion píldora formativa Thoth nº 17, ¿Qué es la cifra afín?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora017.pdf

Lectura recomendada (2/2)

- Introducción a la seguridad informática y criptografía clásica, Lección 8 Algoritmos de cifra por sustitución monoalfabética, MOOC Crypt4you, Jorge Ramió, 2016
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion8.html
- Libro electrónico de seguridad informática y criptografía, versión 4.1, Jorge Ramió, 2006, archivo Criptografía clásica (Word y PDF), 105 páginas
  - https://www.criptored.es/guiateoria/gt_m001a.htm

Class4crypt c4c6.4

Módulo 6. Algoritmos de criptografía clásica

Lección 6.4. Criptoanálisis a la sustitución monoalfabética

6.4.1. Al-Kindi y las estadísticas del lenguaje

6.4.2. Criptoanálisis a la cifra por desplazamiento puro

6.4.3. Criptoanálisis a la cifra por decimación pura

6.4.4. Criptoanálisis a la cifra afín

6.4.5. Mejorando la seguridad de la cifra con sustitución polialfabética

Class4crypt c4c6.4 Criptoanálisis a la sustitución monoalfabética
https://www.youtube.com/watch?v=ZrmeAqRMfO8

El aporte de Al-Kindi

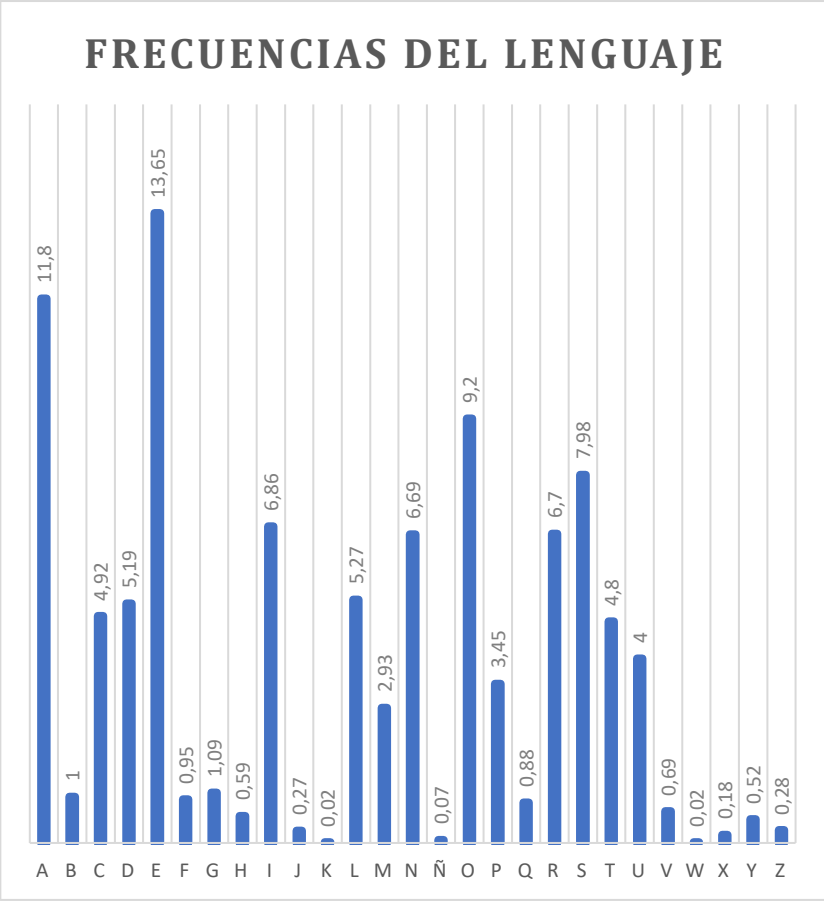
- En el siglo IX, el filósofo y matemático Al-Kindi da las pautas a seguir para criptoanalizar un cifrado por sustitución monoalfabética en su manuscrito “Descifrando mensajes criptográficos”
  - "Una forma de resolver un mensaje cifrado, si conocemos su idioma, es encontrar un texto en claro diferente del mismo idioma, suficientemente largo como para llenar una hoja más o menos, y luego contamos las ocurrencias de cada letra. Llamaremos a la letra que aparece con mayor frecuencia la 'primera', la siguiente letra más frecuente será “segunda” … el símbolo que más aparece en el criptograma lo cambiamos por esa 'primera' letra del texto en claro …”

Estadísticas y redundancia del lenguaje

- El lenguaje castellano módulo 27 tiene una gran redundancia
- Una cifra monoalfabética no cambia los valores de frecuencia en el criptograma

Valores de frecuencia relativa de un texto mod 27 expresados en tanto por ciento					
Frecuencia Alta		Frecuencia Media		Frecuencia Baja	
E	13,11	C	4,85	Y	0,79
A	10,60	L	4,42	Q	0,74
S	8,47	U	4,34	H	0,60
O	8,23	M	3,11	Z	0,26
I	7,16	P	2,71	J	0,25
N	7,14	G	1,40	X	0,15
R	6,95	B	1,16	W	0,12
D	5,87	F	1,13	K	0,11
T	5,40	V	0,82	Ñ	0,10

ESTIRANDO 73%



Criptoanálisis al desplazamiento puro

- $c_i = m_i + b \bmod 27$
- C = JXYJJ XUFWF QNJPU FNXFÑ JQFXM JWQTX TDJPQ FXYWN XYJIJ PQZRI TJXJP QNXQT UFNXF ÑJIJP FUFLN RFFRY JWNTW VZJMJ INGZÑ FITZR FAJEQ FXUFW FVZJP TAJFR GNJRK ZJFVZ NITRI JJPUW NRHNU NYTFU FWJHN TXTGW JPFYN JWWFI JXFUF WJHNJ RITPZ JLT
- Las letras más frecuentes son la J con un 16,0%, F con 13,3% y N con 8,5%
- ¿Será J el resultado de cifrar la letra E?

Frecuencias observadas en el criptograma

Carácter	Frecuencia	Num Apa...
J	15,957%	30
F	13,298%	25
N	8,511%	16
T	6,915%	13
X	6,915%	13
W	6,383%	12
R	4,787%	9
I	4,787%	9
P	4,787%	9
U	4,787%	9
Z	4,255%	8
Q	4,255%	8
Y	3,191%	6
H	1,596%	3
G	1,596%	3
V	1,596%	3
Ñ	1,596%	3
A	1,064%	2
L	1,064%	2
M	1,064%	2
K	0,532%	1
D	0,532%	1
E	0,532%	1

Frecuencias del alfabeto Español

Carácter	Frecuencia
E	13.65%
A	11.797%
O	9.195%
S	7.983%
R	6.696%
N	6.69%
I	6.86%
L	5.27%
D	5.19%
C	4.919%
T	4.802%
P	3.445%
U	3.996%
M	2.925%
B	1.0%
F	0.953%
V	0.693%
Q	0.875%
G	1.043%
H	0.585%
J	0.272%
Ñ	0.074%
K	0.022%
W	0.019%
X	0.183%
Y	0.523%
Z	0.291%

Resolviendo ecuaciones mod 27

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

- Si suponemos que la letra E, la más frecuente del castellano, se ha cifrado como la letra J que es la que tiene una mayor frecuencia en el criptograma:
- Cifrado: $J = E + b \mod 27$ $9 = 4 + b \mod 27$ $b = 9 - 4 \mod 27 = 5$
- Cifrado: $c_i = m_i + 5 \mod 27$ Descifrado: $m_i = c_i + 22 \mod 27$
- Descifrando primer bloque del criptograma C = JXYJJ
  - $m_1 = J + 22 \mod 27 = 9 + 22 \mod 27 = 31 \mod 27 = 4 = E$
  - $m_2 = X + 22 \mod 27 = 24 + 22 \mod 27 = 46 \mod 27 = 19 = S$
  - $m_3 = Y + 22 \mod 27 = 25 + 22 \mod 27 = 47 \mod 27 = 20 = T$
  - $m_4 = J + 22 \mod 27 = 9 + 22 \mod 27 = 31 \mod 27 = 4 = E$
  - $m_5 = J + 22 \mod 27 = 9 + 22 \mod 27 = 31 \mod 27 = 4 = E$

Solución al criptoanálisis

- Texto en claro del ejemplo:
- M = ESTE ES PARA MI EL PAISAJE MAS HERMOSO Y EL MAS TRISTE DEL MUNDO ES EL MISMO PAISAJE DE LA PAGINA ANTERIOR QUE HE DIBUJADO UNA VEZ MAS PARA QUE LO VEAN BIEN FUE AQUÍ DONDE EL PRINCIPITO APARECIO SOBRE LA TIERRA DESAPARECIENDO LUEGO
- “El principito” de Antoine de Saint-Exupéry (1943)
- Pero, a veces, las estadísticas nos pueden dar alguna sorpresa
- Al cifrar el poema “Piececitos de niños” (1924) de Gabriela Mistral con 76 palabras y 360 letras, nos encontramos que ni la letra E ni la letra A son las que muestran una mayor frecuencia en el poema, sino la letra O
- Ejercicio práctico con el software Criptoclásicos v2.1

Cifrado por sustitución con b = 9

Piececitos de niño,
azulosos de frío,
¡cómo os ven y no os cubren,
Dios mío!
¡Piececitos heridos
por los guijarros todos,
ultrajados de nieves
y lodos!
El hombre ciego ignora
que por donde pasáis,
una flor de luz viva
dejáis;
que allí donde ponéis
la plantita sangrante,
el nardo nace más
fragante.
Sed, puesto que marcháis
por los caminos rectos,
heroicos como sois
perfectos.
Piececitos de niño,
dos joyitas sufrientes,
¡cómo pasan sin veros
las gentes!

C = YQNLN LQCXB MNVQW XJIDT
XBXBM NÑAQX LXUXX BENVH
VXXBL DKANV MQXBU QXYQN
LNLQC XBPNA QMXBY XATXB
ODQRJ AAXBC XMXBD TCAJR
JMXBM NVQNE NBHTX MXBNT
PXUKA NLQNO XQOVX AJZDN
YXAMX VMNYJ BJQBD VJÑTX
AMNTD IEQEJ MNRJQ BZDNJ
TTQMX VMNYX VNQBT JYTJV
CQCJB JVOAJ VCNNT VJAMX VJLNU
JBÑAJ OJVCN BNMYD NBCXZ
DNUJA LPJQB YXATX BLJUQ
VXBAN LCXBP NAXQL XBLXU
XBXQB YNAÑN LCXBY QNLNL
QCXBM NVQWX MXBRX HQCJB
BDÑAQ NVCNB LXUXY JBJVB
QVENA XBTJB ONVCN

Criptoanálisis a la decimación pura

- $c_i = a * m_i \bmod 27$
- C = FHHMI MTAJY KMPMJ JKFWX MAYMX AÑFHM PKXAX TMJFH PAOKW MBGFF JAOAQ PMJYM LHAWK EAQFN GFTMJ FHOAQ HAJUM FWYJF YFWKF WXMQF FWPMU FJAÑF HHAWA QFWPM JJFJY JAQHA QOAJK TMQAQ LFWCA PFJJA OMQPM WHAQN HMJFP KHHAQ BGFFW PMWYJ AIATM PMYAJ XMFHH MIMFW HHFUA JAPAQ AXFHA AIGFH AUMHT FAYMP YMP
- Las letras más frecuentes son la A con un 14,1%, F con 12,9% y M con 12,5%
- ¿Será F el resultado de cifrar la letra E?

Frecuencias observadas en el criptograma

Carácter	Frecuencia	Num Apa...
A	14,163%	33
F	12,876%	30
M	12,446%	29
J	9,013%	21
H	9,013%	21
W	6,009%	14
P	6,009%	14
Q	5,579%	13
Y	4,292%	10
K	3,433%	8
X	3,004%	7
T	2,575%	6
O	2,146%	5
U	1,717%	4
G	1,717%	4
I	1,717%	4
N	0,858%	2
L	0,858%	2
B	0,858%	2
Ñ	0,858%	2
C	0,429%	1
E	0,429%	1

Frecuencias del alfabeto Español

Carácter	Frecuencia
E	13.65%
A	11.797%
O	9.195%
S	7.983%
R	6.696%
N	6.69%
I	6.86%
L	5.27%
D	5.19%
C	4.919%
T	4.802%
P	3.445%
U	3.996%
M	2.925%
B	1.0%
F	0.953%
V	0.693%
Q	0.875%
G	1.043%
H	0.585%
J	0.272%
Ñ	0.074%
K	0.022%
W	0.019%
X	0.183%
Y	0.523%
Z	0.291%

Resolviendo ecuaciones mod 27

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

- Si suponemos que la letra E, la más frecuente del castellano, se ha cifrado como la letra F que es la segunda más frecuente en el criptograma:
- Cifrado: $F = a * E \mod 27$ $5 = a * 4 \mod 27$
- $a = 5 * \text{inv}(4, 27) = 5 * 7 \mod 27 = 8$ $\text{inv}(8, 27) = 17$
- Cifrado: $c_i = 8 * m_i \mod 27$ Descifrado: $m_i = c_i * 17 \mod 27$
- Descifrando las primeras 6 letras del criptograma FHHMI M
  - $m_1 = F * 17 \mod 27 = 5 * 17 \mod 27 = 85 \mod 27 = 4 = E$
  - $m_2 = H * 17 \mod 27 = 7 * 17 \mod 27 = 119 \mod 27 = 11 = L$
  - $m_3 = H * 17 \mod 27 = 7 * 17 \mod 27 = 119 \mod 27 = 11 = L$

Solución al criptoanálisis

- m_4 = M * 17 mod 27 = 12 * 17 mod 27 = 204 mod 27 = 15 = O
- m_5 = I * 17 mod 27 = 8 * 17 mod 27 = 136 mod 27 = 1 = B
- m_6 = M * 17 mod 27 = 12 * 17 mod 27 = 204 mod 27 = 15 = O

- Texto en claro del ejemplo:
- M = EL LOBO PARTIO CORRIENDO A TODA VELOCIDAD POR EL CAMINO QUE ERA MAS CORTO Y LA NIÑA SE FUE POR EL MAS LARGO ENTRETENIENDOSE EN COGER AVELLANAS EN CORRER TRAS LAS MARIPOSAS Y EN HACER RAMOS CON LAS FLORECILLAS QUE ENCONTRABA POCO TARDO EL LOBO EN LLEGAR A CASA DE LA ABUELA GOLPEA TOC TOC
- “Caperucita roja” de Charles Perrault (1697)
- Ejercicio práctico con el software Criptoclásicos v2.1

Criptoanálisis a la cifra afín

- $c_i = a * m_i + b \bmod 27$
- C = VNAFL FLVYA FNYMR RIGÑV NJVFR FSTVÑ JFLVR FYAVN YIÑVR AFÑJV RINLV RVNIÑ CVÑIY JÑICM ÑFPFL VYLVV RCTÑI VÑFTN ILVVY IYLMP TWIYÑ VFRMU FLIYL VJFRC FNVÑF DTVRI YIWIY RVYMQ TVNFT NIFLI NLVDT MVÑFD TVVYJ VVRQÑ FNBVÑ CFNIJ VEMQM RFLVA MFNRF YSFRF PÑFYF RSMV
- Las letras más frecuentes son la V con un 16,3%, F con 13,4% e I con 8,1%
- ¿Será V el resultado de cifrar la letra E y F el resultado de cifrar la letra A?

Frecuencias observadas en el criptograma

Carácter	Frecuencia	Num Apa...
V	16,268%	34
F	13,397%	28
I	8,134%	17
R	8,134%	17
Ñ	7,656%	16
Y	7,656%	16
N	6,699%	14
L	6,220%	13
M	4,785%	10
T	4,306%	9
J	3,349%	7
C	2,392%	5
A	2,392%	5
Q	1,435%	3
D	1,435%	3
P	1,435%	3
S	1,435%	3
W	0,957%	2
E	0,478%	1
B	0,478%	1
G	0,478%	1
U	0,478%	1

Frecuencias del alfabeto Español

Carácter	Frecuencia
E	13.68%
A	12.53%
O	8.68%
S	7.98%
R	6.87%
N	6.71%
I	6.25%
D	5.86%
C	5.68%
L	4.97%
T	4.63%
U	3.93%
M	3.15%
P	2.51%
B	1.42%
G	1.01%
V	0.9%
Q	0.88%
F	0.69%
Ñ	0.074%
J	0.44%
H	0.7%
W	0.02%
K	0.01%
X	0.22%
Y	0.9%
Z	0.52%

Resolviendo ecuaciones mod 27

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

- Si suponemos que la letra E, la más frecuente del castellano, se ha cifrado como la letra V y que la letra A, la segunda más frecuente del castellano, se ha cifrado como F, planteamos dos ecuaciones independientes:
- Ecuación 1: $V = a * E + b \bmod 27$ $\quad$ $22 = a * 4 + b \bmod 27$
- Ecuación 2: $F = a * A + b \bmod 27$ $\quad$ $5 = a * 0 + b \bmod 27$ $\quad$ $(b = 5)$
- Reemplazando b en ecuación 1: $\quad$ $22 = a * 4 + 5 \bmod 27$
- $a = (22 - 5) * \text{inv}(4, 27) \bmod 27 = 17 * 7 \bmod 27 = 119 \bmod 27 = 11$
- Ecuación cifrado: $c_i = 11 * m_i + 5 \bmod 27$
- Ecuación descifrado: $m_i = (c_i - 5) * \text{inv}(11, 27) \bmod 27 = (c_i - 5) * 5 \bmod 27$

Solución al criptoanálisis

- Descifrando el primer bloque VNAFL
  - $m_1 = (V - 5) * 5 \bmod 27 = (22 - 5) * 5 \bmod 27 = 85 \bmod 27 = 4 = E$
  - $m_2 = (N - 5) * 5 \bmod 27 = (13 - 5) * 5 \bmod 27 = 40 \bmod 27 = 13 = N$
  - $m_3 = (A - 5) * 5 \bmod 27 = (0 - 5) * 5 \bmod 27 = -25 \bmod 27 = 2 = C$
  - $m_4 = (F - 5) * 5 \bmod 27 = (5 - 5) * 5 \bmod 27 = 0 \bmod 27 = 0 = A$
  - $m_5 = (L - 5) * 5 \bmod 27 = (11 - 5) * 5 \bmod 27 = 30 \bmod 27 = 3 = D$
- M = EN CADA DESCANSILLO FRENTE A LA PUERTA DEL ASCENSOR EL CARTELON DEL ENORME ROSTRO MIRABA DESDE EL MURO ERA UNO DE ESOS DIBUJOS REALIZADOS DE TAL MANERA QUE LOS OJOS LE SIGUEN A UNO ADONDE QUIERA QUE ESTE EL GRAN HERMANO TE VIGILA DECIAN LAS PALABRAS AL PIE
- “1984” de George Orwell (1949) → Criptoclásicos

Necesidad de una cifra polialfabética

- Va a enmascarar mejor la redundancia del lenguaje en el criptograma
- Se inicia en 1466 por Leon Battista Alberti
- Par de discos, interior y exterior, que giran de forma independiente, permitiendo cambiar la letra cifrada (clave)
- Blaise de Vigenère será su máximo exponente

Más información en píldoras Thoth

Píldora 18: ¿Cómo se ataca la cifra por sustitución monoalfabética?

https://www.youtube.com/watch?v=Jsg4QNvhOYQ

Conclusiones de la Lección 6.4

- Al-Kindi nos muestra cómo atacar la cifra por sustitución monoalfabética en el siglo IX, haciendo coincidir letras de muy alta frecuencia de un texto en claro con aquellas letras de alta frecuencia que aparecen en el criptograma
- Esto es debido a la redundancia del lenguaje, que se manifiesta de forma muy clara en el criptograma en este tipo de cifrados monoalfabéticos
- En el caso de tener dos incógnitas, como el factor a de decimación y el factor b de desplazamiento en la cifra afín, será necesario plantear dos ecuaciones independientes, también siguiendo las estadísticas del lenguaje
- Lo normal es que las letras más frecuentes del criptograma se deban al cifrado de las letras más frecuentes del castellano, primero la letra E, después la letra A y luego la letra O, aunque puede haber variaciones entre ellas

Lectura recomendada

- The Breakthrough of Frequency Analysis, University of Klagenfurt, Austria
  - http://cs-exhibitions.uni-klu.ac.at/index.php?id=279
- Guion píldora formativa Thoth nº 18, ¿Cómo se ataca la cifra por sustitución monoalfabética?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora018.pdf
- Introducción a la seguridad informática y criptografía clásica, Lección 8 Algoritmos de cifra por sustitución monoalfabética, MOOC Crypt4you, Jorge Ramió, 2016
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion8.html
- Libro electrónico de seguridad informática y criptografía, versión 4.1, Jorge Ramió, 2006, archivo Criptografía clásica (Word y PDF), 105 páginas
  - https://www.criptored.es/guiateoria/gt_m001a.htm

Class4crypt c4c6.5

Módulo 6. Algoritmos de criptografía clásica

Lección 6.5. Sustitución polialfabética y algoritmo de Vigenère

6.5.1. La cifra por sustitución polialfabética soluciona la vulnerabilidad de la monoalfabética

6.5.2. La figura de Blaise de Vigenère

6.5.3. El cifrado y la tabla de Vigenère

6.5.4. Característica de la cifra de Vigenère

6.5.5. Ejercicios prácticos de cifrado y descifrado de Vigenère

Class4crypt c4c6.5 Sustitución polialfabética y algoritmo de Vigenère
https://www.youtube.com/watch?v=dEiHh4lA6Z4

Sustitución monoalfabética no era segura

- A B C D E F G H I J K L M N Ñ O P Q R S T U V W X Y Z
- 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
- Los algoritmos de cifra clásica por sustitución monoalfabética, como por ejemplo $c_i = m_i + b \bmod 27$, no eran seguros, es muy fácil romperlos
- Si b = 3 (César), como la letra D tiene el código 3, seria igual cifrar así:
- **E**ST**E** **E**S UN **E**J**E**MPLO D**E** CIFRA POR SUSTITUCIÓN D**E**L TIPO C**E**SAR **+**
- DDDD DD DD DDDDDDD DD DDDDD DDD DDDDDDDDDDD DDD DDDD DDDDD **mod 27**
- **H**VW**H** **H**V XP **H**M**H**OSÑR G**H** FLIUD SRU VXVWLWXFLRP G**H**Ñ WLSR F**H**VDU
- La redundancia del lenguaje se manifiesta de manera clara en el criptograma, incluso con la cifra afín ($c_i = a*m_i + b \bmod 27$). En histograma de frecuencias de las letras del criptograma, sólo se desplazarán las “marcas” de esas letras

Frecuencias del lenguaje y criptogramas

$c_i = m_i + b \bmod 27$
El desplazamiento es en bloque, es simple observarlo…

$c_i = a*m_i + b \bmod 27$
El desplazamiento ya se hace letra a letra, pero las frecuencias son las mismas…

Solución

Para obtener algo así en el criptograma, cambiaremos la D anterior por una clave con letras diferentes y logramos así difuminar la redundancia del lenguaje en el criptograma

Blaise de Vigenère

- Blaise de Vigenère fue un criptógrafo, químico y diplomático francés: nace en 1523 y fallece en 1596
- En su libro *Traité des chiffres où secrètes manières d'escrire* publicado en 1586 describe de una forma detallada y estructurada la cifra polialfabética, basándose en trabajos previos de importantes criptógrafos como Alberti, Trithemius, Bellaso y della Porta
- Por este motivo, el sistema de cifra por sustitución polialfabética más conocido y estudiado durante muchos siglos, lleva su nombre: Vigenère
- Su cifra fue invulnerable durante más de 275 años

Cifrado de Vigenère

- Al usar una clave, se soluciona la debilidad del cifrado del César en que una letra se cifra siempre igual (+3 = D)
- La cifra se hace letra a letra (monográmica), igual que en César
- En aquellos años se usaba una tabla para la sustitución de las letras del texto en claro por otras en el criptograma. Hoy en día se usa matemáticas discretas
- Se construye una tabla que permitirá el cifrado y el descifrado tal y como se muestra en la diapositiva siguiente. La primera fila de la tabla marcará la letra del texto en claro y la primera columna de la tabla marcará la letra de la clave
- Cada letra del texto en claro se reemplaza por la letra que se encuentra en la intersección entre la columna del texto en claro y la fila de la clave
- Si el texto en claro es M = HOLA AMIGOS y la clave es K = CIFRA, la H del texto en claro se cifrará con la C de la clave, la O del texto con la I de la clave, etc.

Tabla de Vigenère

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

Texto	Clave	Criptograma
H	C	J
O	I	W
L	F	P
A	R	R
A	A	A
M	C	Ñ
I	I	P
G	F	L
O	R	G
S	A	A

Criptograma: JWPRA ÑPLGS

Características de la cifra de Vigenère

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

Clave: 0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

M	**H**	**O**	**L**	**A**	**A**	**M**	**I**	**G**	**O**	**S**
K	C	I	F	R	A	C	I	F	R	A
	7+2	15+8	11+5	0+18	0+0	12+2	8+8	6+5	15+18	19+0
	9	23	16	18	0	14	16	11	33	19
	9	23	16	18	0	14	16	11	6	19
	J	**W**	**P**	R	A	Ñ	**P**	L	**G**	S

Se cifra letra a letra, sumando módulo 27 el valor o código del texto en claro con los valores de esta clave

$$c_i = m_i + k_i \bmod 27$$

- Dos letras iguales del texto en claro (O) pueden cifrarse de forma diferente, en un caso con la clave I (dando W) y en el otro caso con la clave R (dando G)
- Dos letras iguales del criptograma (P) pueden provenir de letras diferentes

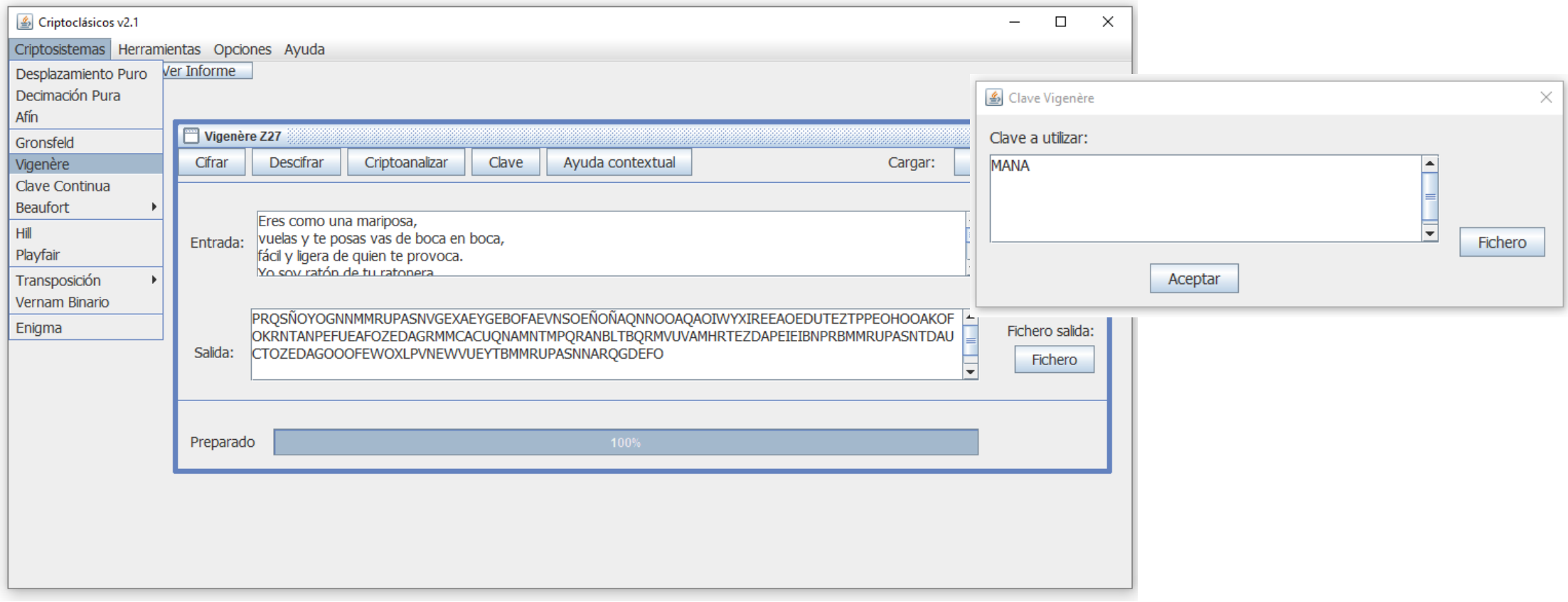
Ejercicio práctico 1: cifrado Vigenère

- Cifra con Vigenère y la clave MANÁ el texto que se indica con Criptoclásicos v2.1: https://www.criptored.es/software/sw_m001c.htm

Eres como una mariposa,
vuelas y te posas vas de boca en boca,
fácil y ligera de quien te provoca.
Yo soy ratón de tu ratonera,
trampa que no mata pero no libera,
vivo muriendo prisionero.
Mariposa traicionera,
todo se lo lleva el viento,
mariposa no regreso.

https://www.youtube.com/watch?v=av3wkasS-WQ

Resultado del cifrado de Vigenère



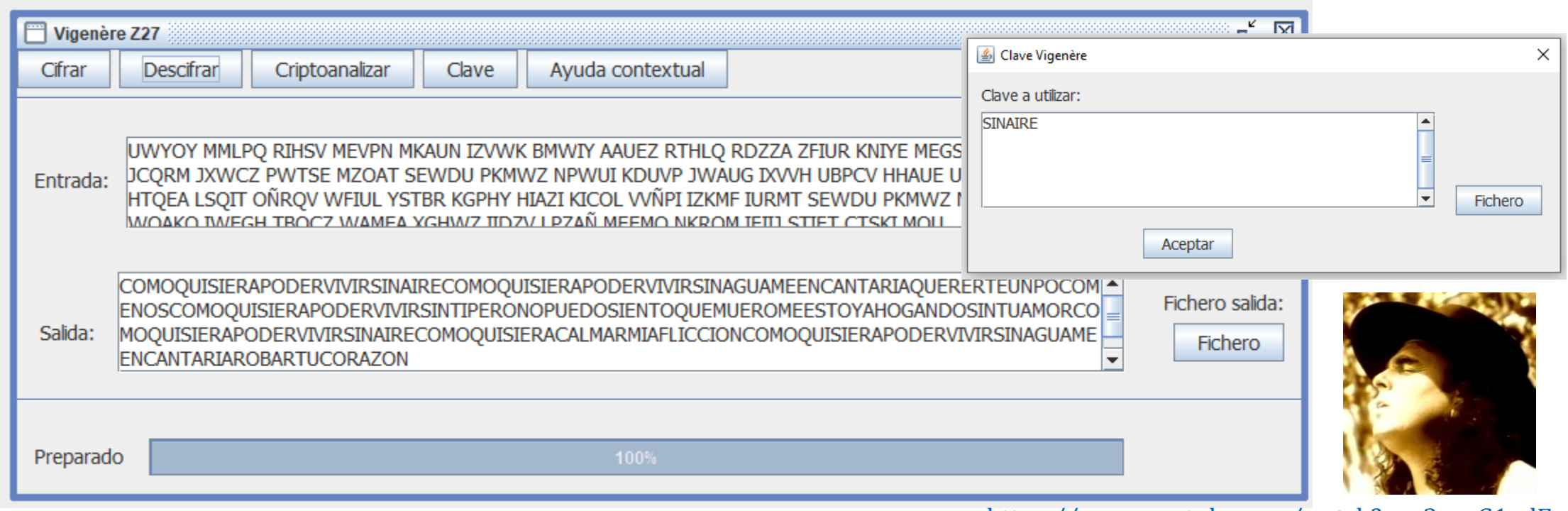
C = PRQSÑ OYOGN NMMRU PASNV GEXAE YGEBO FAEVN SOEÑO ÑAQNN OOAQA OIWYX IREEA OEDUT EZTPP EOHOO AKOFO KRNTA NPEFU EAFOZ EDAGR MMCAC UQNAM NTMPQ RANBL TBQRM VUVAM HRTEZ DAPEI EIBNP RBMMR UPASN TDAUC TOZED AGOOO FEWOX LPVNE WVUEY TBMMR UPASN NARQG DEFO

Ejercicio práctico 2: descifrado Vigenère

- Con Criptoclásicos v2.1 descifra con Vigenère y clave SIN AIRE el criptograma que se indica: https://www.criptored.es/software/sw_m001c.htm

C = UWYOY MMLPQ RIHSV MEVPN MKAUN IZVWK BMWIY AAUEZ RTHLQ RDZZA ZFIUR KNIYE MEGSU GAZZE JCQRM JXWCZ PWTSE MZOAT SEWDU PKMWZ NPWUI KDUVP JWAUG IXVVH UBPCV HHAUE ULSJC QMCVV HTQEA LSQIT OÑRQV WFIUL YSTBR KGPHY HIAZI KICOL VVÑPI IZKMF IURMT SEWDU PKMWZ NCICP SZYII WOAKO IWEGH TBQCZ WAMEA XGHWZ IIDZV LPZAÑ MEEMQ NKRQM IEIIJ STIET CTSKI MOU

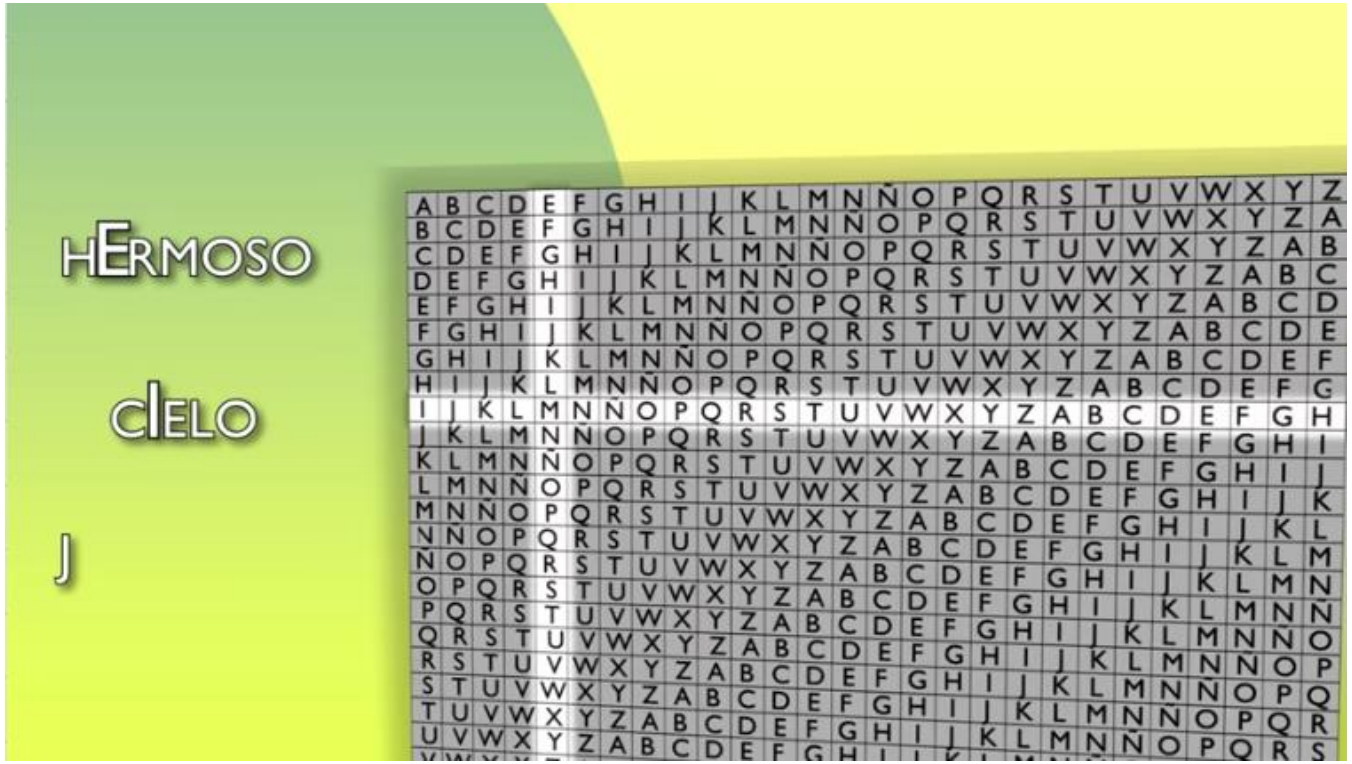
Resultado del descifrado de Vigenère



https://www.youtube.com/watch?v=g3uxeG1rrlE

M = COMO QUISIERA PODER VIVIR SIN AIRE, COMO QUISIERA PODER VIVIR SIN AGUA. ME ENCANTARIA QUERERTE UN POCO MENOS, COMO QUISIERA PODER VIVIR SIN TI. PERO NO PUEDO, SIENTO QUE MUERO, ME ESTOY AHOGANDO SIN TU AMOR. COMO QUISIERA PODER VIVIR SIN AIRE, COMO QUISIERA CALMAR MI AFLICCION. COMO QUISIERA PODER VIVIR SIN AGUA, ME ENCANTARIA ROBAR TU CORAZON.

Más información en píldoras Thoth



https://www.youtube.com/watch?v=8Pb5U64iwV4

Conclusiones de la lección 6.5

- La cifra por sustitución monoalfabética, utilizada desde el siglo I antes de Cristo por el sistema del César, es insegura porque la redundancia del lenguaje se manifiesta de forma muy clara en el criptograma
- La cifra por sustitución polialfabética, en particular el algoritmo de Vigenère, soluciona este problema al difuminar las frecuencias de las letras en el criptograma, aumentando la frecuencia de las letra poco frecuentes (K, Q, X) y disminuyendo la frecuencia de las letras muy frecuentes (A, E, O)
- Una misma letra del texto en claro entregará diferentes criptogramas al ser cifrada por una parte o letra diferente de la clave. Además, letras iguales en el criptograma pueden provenir de diferentes letras del textos en claro
- La cifra de Vigenère inventada en 1586, se criptoanaliza en 1863 mediante el método de Kasiski: fue invulnerable durante 277 años

Lectura recomendada

- Introducción a la seguridad informática y criptografía clásica. Lección 9: Algoritmos de cifra por sustitución polialfabética, MOOC Crypt4you, Jorge Ramió, 2017
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion9.html
- Guion píldora formativa Thoth nº 19, ¿Qué es la cifra de Vigenère?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora019.pdf
- Cuadernos de Laboratorio de Criptografía CLCript 03: Cifrado, descifrado y criptoanálisis de Vigenère, Jorge Ramió, 2019
  - https://www.criptored.es/descarga/CLCript_entrega_03_Cifrado_descifrado_criptoanalisis_Vigenere.pdf
- Vigenère cipher, Wikipedia
  - https://en.wikipedia.org/wiki/Vigen%C3%A8re_cipher

Class4crypt c4c6.6

Módulo 6. Algoritmos de criptografía clásica

Lección 6.6. Criptoanálisis a la cifra de Vigenère por el método Kasiski

6.6.1. Recordando la cifra por sustitución polialfabética periódica de Vigenère

6.6.2. La redundancia del lenguaje

6.6.3. La figura de Friedrich Kasiski y otros criptólogos de la época

6.6.4. Desarrollo del método de Kasiski para el criptoanálisis de Vigenère

6.6.5. Ejercicio práctico de criptoanálisis a Vigenère por el método de Kasiski

Class4crypt c4c6.6 Criptoanálisis a la cifra de Vigenère por el método Kasiski
https://www.youtube.com/watch?v=K3tpKeDQs6s

Recordando la cifra de Vigenère

- Sea la clave MAR y el texto a cifrar el poema “El mar” de Pablo Neruda

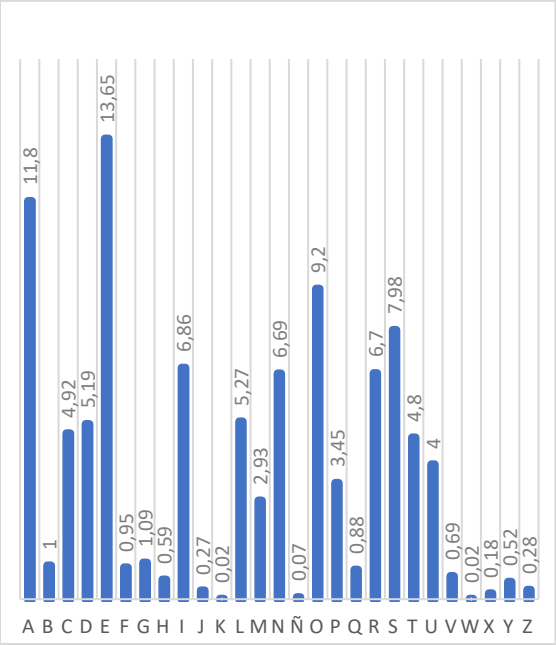
Necesito del mar porque me enseña:
no sé si aprendo música o conciencia:
no sé si es ola sola o ser profundo
o sólo ronca voz o deslumbrante
suposición de peces y navíos.

Redundancia del lenguaje

```
M = NEC ESI TOD ELM ARP ORQ UEM EEN SEÑ ANO SES IAP REN DOM USI CAO CON CIE NCI ANO SES IES OLA
K = MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR
C = YET PSZ FOU PLD MRH ARI GED PEE EEF MNG EEK TAH DEE OOD GSZ ÑAG ÑOE ÑIV YCZ MNG EEK TEK ALR

M = SOL AOS ERP ROF UND OOS OLO RON CAV OZO DES LUM BRA NTE SUP OSI CIO NDE PEC ESY NAV IOS
K = MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR MAR
C = EOC MOK PRH DOW GNU AOK ALG DOE ÑAN AZG OEK WUD NRR YTV EUH ASZ ÑIG YDV BET PSP YAN TOK
```

La redundancia del lenguaje



- Todos los lenguajes son redundantes:
- Letras (monogramas) más frecuentes que otras: E vs Q
- Dos letras (digramas) más frecuentes que otras: DE vs QA
- Tres letras (trigramas) más frecuentes que otras: ADO vs FXE
- Cuatro letras (tetragramas) más frecuentes que otras: ENTE vs TAZC
- Incluso de cinco letras… MENTE, IENDO…
- Y esos bloques de letras podrían cifrarse con la misma parte de la clave, dando como resultado el mismo criptograma
- Por lo tanto, bloques de 3 o más letras repetidas en un criptograma NO se darán por simple casualidad. No así si esos bloques son de sólo dos letras
- En esta propiedad del lenguaje se basó Kasiski para romper la cifra de Vigenère

Profundizando en esa redundancia

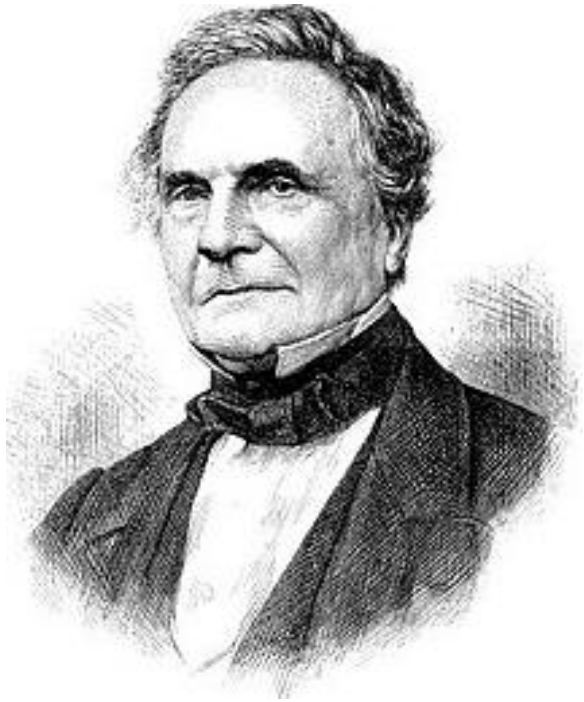
- Texto en claro a cifrar con Vigenère (con 18 tetragramas ENTE)
  - El agente silente de Occidente, seguirá al frente de los puntos calientes en el Extremo Oriente, mientras allí haya gente que aliente que este valiente siga allí presente, que no ausente. Aunque asiente con entereza que sea de forma aparente. Atentamente, el Teniente Vicente Morente.
  - Clave: BOND
- ELAG **ENTE** SIL **ENTE** DEOCCID **ENTE** SEGUIRAALFR **ENTE** DELOSPUNTOSCALI **ENTE** SENELEXTREMOORI **ENTE** MIENTRASALLIHAYAG **ENTE** QUEALI **ENTE** QUEESTEVALI **ENTE** SIGAALLIPRES **ENTE** QUENOAUS **ENTE** AUNQUEASI **ENTE** CON **ENTE** REZAQUESEADEFORMAAPAR **ENTE** ATENTAM **ENTE** ELTENI **ENTE** VIC **ENTE** MOR **ENTE**
- FZNJ **FBGH** TWX **HÑIQ** GFDOFJR **QPUS** FHHJUUBOXIS **SZWF** RQÑPHCXÑIBVDOXL **FBGH** TSZHMSKWSSYRPGU **HÑIQ** OJSZWSOFDMZUKBNNJ **FBGH** RJQDMW **QPUS** DXFSFWFKNÑJ **SZWF** HUJBOXÑJEEHT **SZWF** FHHÑDNXT **SZWF** OHPRJQDTW **QPUS** ORÑ **SZWF** GQCBFHHTSNGFTBUNONSBG **QPUS** NWFBGDN **SZWF** SXWFBU **HÑIQ** YJQ **QPUS** YRS **SZWF**

Trabajo previo al método de Kasiski

- En el año 1854, 268 años después de la invención de la cifra de Vigenère, el matemático y criptógrafo inglés Charles Babbage criptoanaliza el cifrado por sustitución polialfabética conocido como Autoclave, mucho más seguro y complejo que Vigenère
- Pero Babbage comete el gran error de no publicar sus resultados
- Cifrado Autoclave:
- Sea el texto en claro M = TE ENVIO UN ABRAZO Y UN BESO
- Y sea la clave K = AMIGO
- Una vez se termina la clave, ésta continúa usándose como clave el mismo texto en claro

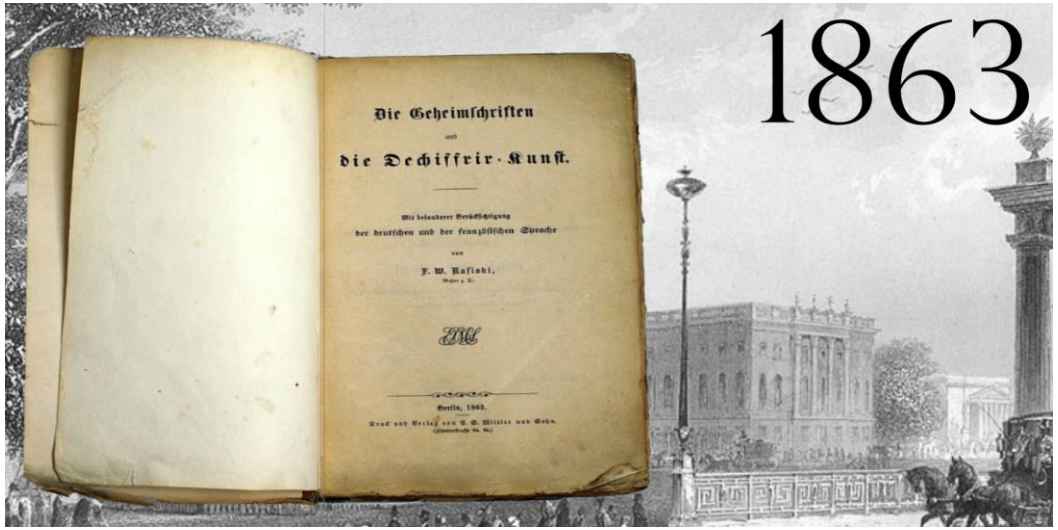
```
        TEENV IOUNA BRAZO YUNBE SO
+       AMIGO TEENV IOUNA BRAZO YU   mod 27
=       TPMSK BSYZV JGUMO ZMNAS QJ
```

- Para saber cómo se ataca esta cifra autoclave, algo más complejo y basado en las estadísticas de los digramas, véase la bibliografía



Friedrich Kasiski

- Friedrich Kasiski fue un militar alemán (1805 - 1881) que en su libro “Die Geheimschriften und die Dechiffrierkunst” (1863), “Escritura secreta y el arte de descifrar”, nos muestra cómo criptoanalizar la cifra polialfabética periódica de Vigenére (1586), un total de 277 años de fortaleza



- "Kasiski muere el 22 de mayo de 1881, sin darse cuenta de que había iniciado una revolución en la criptografía", The Codebreakers, David Kahn, 1996

Reflexiones sobre el ataque de Kasiski

1. Si hay varias cadenas de letras repetidas en el criptograma, entonces el máximo común divisor del espacio que separa a cadenas repetidas nos daré una pista sobre la longitud L de la clave
2. Si esas distancias (letras que las separan) son L_1, L_2, L_3,… L_n, entonces podríamos esperar que L = mcd (L_1, L_2, L_3,… L_n)
3. Si obtenemos un valor L > 1, por ejemplo L = 5, la clave tendría 5 letras
4. Dividimos el criptograma en 5 subcriptogramas C_1, C_2, C_3, C_4 y C_5. Así, C_1 se configurará con las letras en las posiciones 1, 6, 11, 16, 21,… etc., C_2 se configurará con las letras en las posiciones 2, 7, 12, 17, 22,… etc., y C_5 se configurará con las letras en posiciones 5, 10, 15, 20, 25,… etc.
5. Cada subcriptograma Ci será entonces una cifra monoalfabética
6. ¿Se sigue manifestando aquí la redundancia del lenguaje? Sí ✓

Los 8 pasos del criptoanálisis de Kasiski

1. Buscar en el criptograma repeticiones de al menos 3 caracteres y anotar la distancia que separa a todas esas repeticiones
2. Encontrar el máximo común divisor de todas esas separaciones. El mcd nos indicará la posible longitud L de la clave
3. Se divide el criptograma en L subcriptogramas tomando las letras de L en L espacios
4. Para cada uno de los L subcriptogramas, se apunta la frecuencia de aparición de cada letra.
5. Se busca en cada uno de los L subcriptogramas las cuatro frecuencias más altas y que, además, cumplan con la distancia que separa a las letras con mayor frecuencia del alfabeto español mod 27, es decir la A, la E, la O y la S. Esto es, que los espacios entre ellas cumplan la siguiente distribución, conocida como la regla AEOS
6. Regla AEOS: Letra A -> + 4 = Letra E -> + 11 = Letra O -> + 4 = Letra S
7. Ubicada la posición de la Letra A, que es la relativa a la letra A del texto en claro y cuyo código es igual a 0, se mira con qué letra se ha cifrado, dando así la letra correspondiente de la clave en esa posición. Esto es así porque el código de la letra A es el 0
8. Se repite este proceso con todos los subcriptogramas para obtener la clave buscada

Paso 1 Kasiski: cadenas de letras repetidas

- Sea el criptograma C de 404 letras el siguiente:

```
PBVRQ VICAD SKAÑS DETSJ PSIED BGGMP SLRPW RÑPWY EDSDE ÑDRDP CRCPQ MNPWK
UBZVS FNVRD MTIPW UEQVV CBOVN UEDIF QLONM VNUVR SEIKA ZYEAC EYEDS ETFPH
LBHGU ÑESOM EHLBX VAEEP UÑELI SEUEF WHUNM CLPQP MBRRN BPVIÑ MTIBV VEÑIC
ANSJA MTJOK MDODS ELPWI UFOZM QMVNF OHASE SRJWR SFQCO TWVMB JGRPW VSUEX
INQRS JEUEM GGRBD GNNIL AGSJI DSVSU EEINT GRUEE TFGGM PORDF OGTSS TOSEQ
OÑTGR RYVLP WJIFW XOTGG RPQRR JSKET XRNBL ZETGG NEMUO TXJAT ORVJH RSFHV
NUEJI BCHAS EHEUE UOTIE FFGYA TGGMP IKTBW UEÑEN IEEU
```

- Entre otras repeticiones de letras en el criptograma C, se observan las siguientes cadenas en el criptograma (en negrita y colores)
  - 3 cadenas **GGMP**, separadas por 256 y 104 espacios
  - 2 cadenas **YEDS**, separadas por 72 espacios
  - 2 cadenas **HASE**, separadas por 156 espacios
  - 2 cadenas **VSUE**, separadas por 32 espacios: X INQRS JEUEM GGRBD GNNIL AGSJI DSVSU E

Mientras más largas sean las cadenas, y aparezcan más veces, mayores posibilidades de éxito

Paso 2 Kasiski: posible longitud de la clave

- Tres cadenas GGMP separadas por 256 y 104 espacios; dos cadenas YEDS separadas por 72 espacios; dos cadenas HASE separadas por 156 espacios y dos cadenas VSUE separadas por 32 espacios
- MCD (256, 104, 72, 156, 32) = 4
- Longitud posible de la clave L = 4 (generamos 4 subcriptogramas)
  - C_1 = {Letras que ocupan las posiciones 1, 5, 9, 13, 17, …., 401}
  - C_2 = {Letras que ocupan las posiciones 2, 6, 10, 14, 18, …., 402}
  - C_3 = {Letras que ocupan las posiciones 3, 7, 11, 15, 19, …., 403}
  - C_4 = {Letras que ocupan las posiciones 4, 8, 12, 16, 20 ,…., 404}

Paso 3 Kasiski: frecuencia subcriptogramas

C_1 = PQ**AA**E**E**PDMRÑ**EE**DCNUSRI**E**CNIONS**AA**E**E**TLUOL**A**UI**E**ULMNII**EAA**OOLU
MN**A**RSOMRSIS**E**RN**A**ISIRTMDTOORLIORR**E**N**E**NO**A**VSNI**AE**OF**A**MT**E**I

C_2 = **B**VDÑTS**B**PPPDÑPPP**B**FD**P**Q**B**UFNUEZCDF**B**ÑM**B**EÑSFN**P**B**B**Ñ**B**ÑNMKD**P**F
QFSJFT**B**PUNJM**B**NGDUNUF**P**FSSÑR**P**FT**P**JT**B**TETTJFU**B**SUTFT**P**BÑE

C_3 = VISSSIGSWWSDCQWZNMWVO**E**QMVIY**E**SPH**EE**X**EEE**WMQRPMVISTMSWO
MO**E**WQWJW**E**Q**E**GDISS**E**T**E**GOOS**E**TYWWGQSXLGMXOHH**E**C**EE**IGGIW**EE**

C_4 = RCKDJEGLRYDRRMK**VV**TU**VV**DLWRKEYEHGSH**V**PL**V**HCPR**V**T**V**DJJDEIZ
VHSRC**V**G**V**XRUGGLJ**V**EGEGRGTQG**V**JXGRKRZGUJRR**V**JHHUEYGKUNU

Primer paso: marcar la letra más frecuente cada subcriptograma. Lo normal es que sea el resultado de haber cifrado la A o la E

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
C_A	12	0	2	3	12	1	0	0	11	0	0	5	6	9	1	10	2	1	9	7	4	5	1	0	0	0	0
C_B	0	14	1	6	4	12	1	0	0	4	1	0	3	6	8	0	14	2	1	6	9	7	1	0	0	0	1
C_C	0	0	2	2	18	0	7	3	7	1	0	1	7	1	0	6	2	6	1	12	3	0	4	12	3	2	1
C_D	0	0	3	5	7	0	12	6	1	7	5	4	1	1	0	0	2	1	13	2	3	6	14	1	2	3	2

Paso 4 Kasiski: la regla AEOS en general

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z

- Buscamos en cada subcriptograma C_i las cuatro letras con una alta frecuencia y que cumplan, ADEMÁS, con una distribución en posiciones relativas separadas igual que las letras A, E, O y S, las cuatro letras más frecuentes del lenguaje mod 27

$$\mathbf{A} + 4 = 0 + 4 = 4 = \mathbf{E}$$

$$E + 11 = 4 + 11 = 15 = \mathbf{O}$$

$$O + 4 = 15 + 4 = 19 = \mathbf{S}$$

- Intentamos buscar cómo se ha cifrado la letra A (código 0) del texto en claro
- Donde esté esa posición relativa de la A, nos marcará la letra de la clave

Paso 4.1 Kasiski: regla AEOS en C_1

C_1 = PQAAEPDMRÑEEDCNUSRIECNIONSAAETLUOLAUIEULMNIIEAAOOLU

MNARSOMRSISERNAISIRTMDTOORLIORRENENOAVSNIAEOFAMTEI

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
12	0	2	3	12	1	0	0	11	0	0	5	6	9	1	10	2	1	9	7	4	5	1	0	0	0	0

+4 → +11 → +4 →

- Aquí, algunas de las letras más frecuentes en el criptograma coinciden con las posiciones relativas de las letras A, E, O y S de un texto en claro
- Por lo tanto, las frecuencias (12, 12, 10, 7) marcarían las posiciones de la A, de la E, de la O y de la S, en esta misma secuencia
- Por lo tanto, la primera letra de la clave podría ser la **A**

Paso 4.2 Kasiski: regla AEOS en C_2

C_2 = BVDÑTSBPPPDÑPPPBFDPQBUFNUEZCDFBÑMBEÑSFNPBBÑBÑNMKD

PFQFSJFTBPUNJMBNGDUNUFPFSSÑRPFTPJTBTETTJFUBSUTFTPBÑE

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	14	1	6	4	12	1	0	0	4	1	0	3	6	8	0	14	2	1	6	9	7	1	0	0	0	1

+4 → +11 → +4 →

- Arriba se muestran las posiciones relativas de las letras A, E, O y S
- Por lo tanto, las frecuencias (14, 12, 14, 9) marcarían las posiciones de la A, de la E, de la O y de la S, en esta misma secuencia
- Por lo tanto, la segunda letra de la clave podría ser la **B** (Clave = AB)

Paso 4.3 Kasiski: regla AEOS en C_3

C_3 = VISSSIGSWWSDCQWZNMWVOEQMVIYESPHEEXEEEWMQRPMVISTMSWO
MOEWQWJWEQEGDISSETEGOOSETYWWGQSXLGMXOHHECEEIGGIWEE

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	0	2	2	18	0	7	3	7	1	0	1	7	1	0	6	2	6	1	12	3	0	4	12	3	2	1

+4 → +11 → +4 →

- Arriba se muestran las posiciones relativas de las letras A, E, O y S
- Por lo tanto, las frecuencias (18, 7, 12, 12) marcarían las posiciones de la A, de la E, de la O y de la S, en esta misma secuencia
- Por lo tanto, la tercera letra de la clave podría ser la **E** (Clave = ABE)

Paso 4.4 Kasiski: regla AEOS en C_4

C_4 = RCKDJEGLRYDRRMKVVTUVVDLWRKEYEHGSHVPLVHCPRVTVDJJDEI

ZVHSRCVGVXRUGGLJVEGEGRGTQGVJXGRKRZGUJRRVJHHUEYGKUNU

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
0	0	3	5	7	0	12	6	1	7	5	4	1	1	0	0	2	1	13	2	3	6	14	1	2	3	2

+4 +4 +11

- Arriba se muestran las posiciones relativas de las letras A, E, O y S
- Por lo tanto, las frecuencias (13, 14, 12, 5) marcarían las posiciones de la A, de la E, de la O y de la S, en esta misma secuencia
- Por lo tanto, la cuarta letra de la clave podría ser la **R** (Clave = ABER)

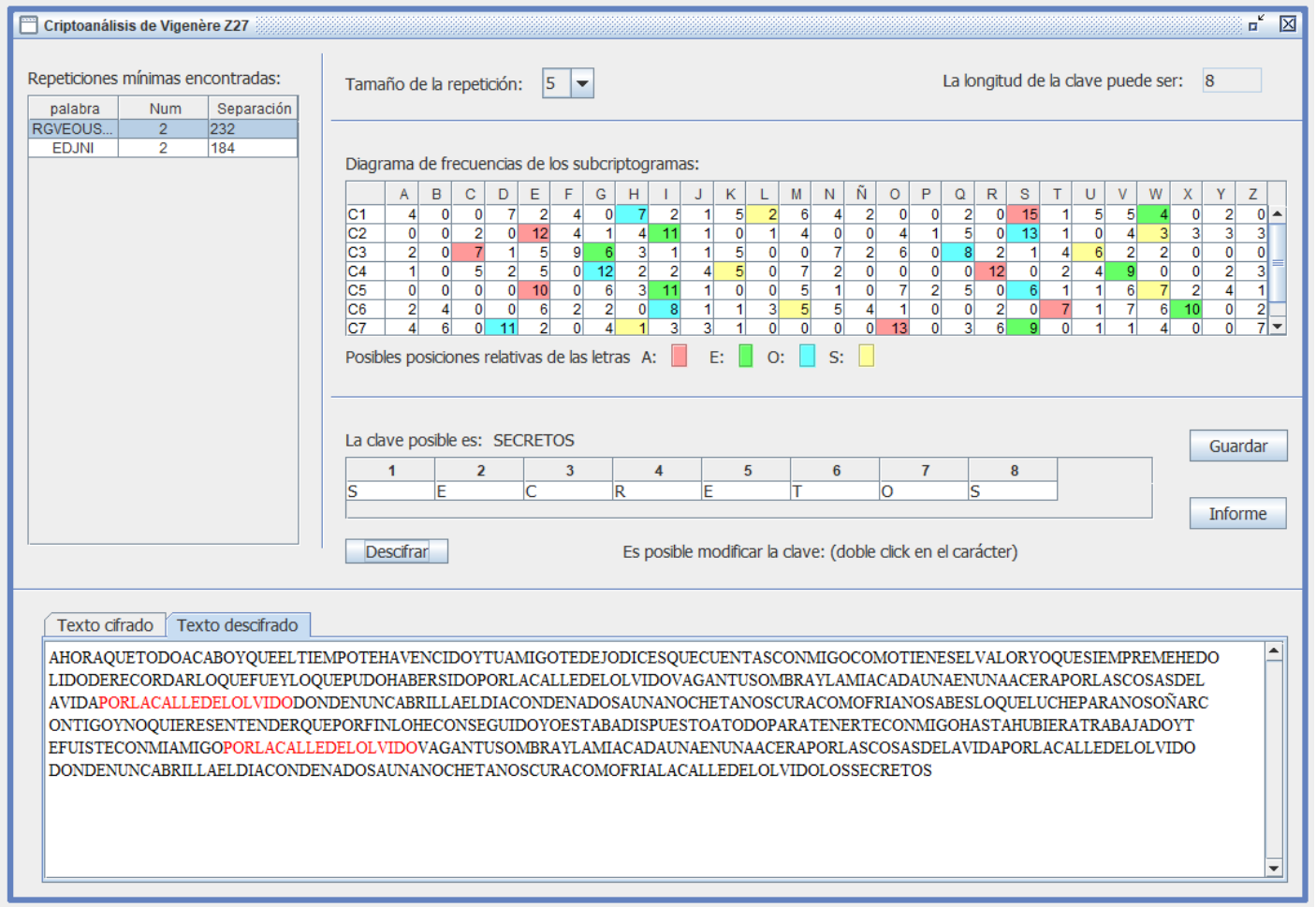
Descifrando el criptograma con clave ABER

- Desciframos ahora el criptograma con la clave ABER, separando los bloques de 5 letras en palabras del castellano y agregando los signos
- PARA QUE LA COSA NO ME SORPRENDA COMO OTROS AÑOS, HE COMENZADO YA CON UNOS SUAVES EJERCICIOS DE PRECALENTAMIENTO. MIENTRAS DESAYUNABA, HE CONTEMPLADO UNA BOLA PLATEADA Y UNA TIRA DE ESPUMILLÓN Y MAÑANA ME INICIARÉ EN EL AMOR AL PRÓJIMO CON LOS QUE LIMPIEN EL PARABRISAS EN LOS SEMÁFOROS. ESTA GIMNASIA DEL CORAZÓN METAFÓRICO ES TAN IMPORTANTE COMO LA DEL OTRO CORAZÓN, PORQUE LOS RIESGOS CORONARIOS ESTAN AHÍ, ESCONDIDOS TRAS LA VIDA SEDENTARIA, Y PARAPETADOS EN FECHAS COMO ESTAS DE NAVIDAD.
- Comienzo del artículo "Gimnasia" del periodista Andrés Aberasturi, publicado el 4/12/94 en el periódico El Mundo (*el año en el que comienzo a impartir la asignatura de Seguridad Informática en la EUI-UPM*)

Ejercicio práctico de ataque por Kasiski

- Criptoanaliza el siguiente texto cifrado con Vigenère de 635 letras, usando Criptoclásicos v2.1: https://www.criptored.es/software/sw_m001c.htm
- SLQJE KJWMS FGEVO THCSM IXZMA IÑHSN SZSZG EGBRH QXWRP BUHMI FVNIR
  AUIUI YXQNW QVRWV DFEMI GGIAH MMGEI MSDÑE NGVRD JNIUZ IFEKW PGYIW
  DDAHQ UILSU HVFRV EDJNI HMIRZ HJYGH YWDZS FGJWB RHIST CEVOD DIFVO
  IZÑAH QNEZO FMYUG PUGSQ OCDMT QSVEW EEXBN FECTI LOIHV NRWVD LSWFV
  OTKAV ERGVE OUSON VHXZH DZKUS WDFVI OMQVO TKMNC EXZVA EEGQW SFSHQ
  KEÑBS FSEYI NOFHW EMVTQ HESHJ MTBHL EDVWE DJNIN MGASI SVCES MDGSV
  EGQNW YHCOG UÑWWK IUVQN SFVIT IYXEH KJKEO IVWUS OKIZJ AVSAG IMIST
  EFZWJ JWLXQ RXIRH IETRX XBWKX GTSGA AYSJR WNOZN FKVVT IKSFC AEWDQ
  MIHMM MIWUS ODMTA AYSRG VEOUS ONVHX ZHDZK USOOY SQVMW IATKE ACEFW
  SUEFR YGOWF YOREV SKSTQ JOTHU HWCKH XZSÑM FRTIG DSGCC OXRWD SNNMW
  DVHQF VQÑBU SFTZO EOWDH KRGIB VWQCU SMONF EOGGA SMSQQ KGÑGS USÑGJ
  LWSDE EROES VWOQC ZBRHD SUKIV GWMSU

Criptoanálisis con Criptoclásicos v2.1



Se encuentra la clave SECRETOS, teniendo cada subcriptograma tan solo 80 letras, y haciendo estadísticas dentro de un grupo de cifra de 27 letras (tres veces menos)

Esto demuestra la gran redundancia que tiene el lenguaje

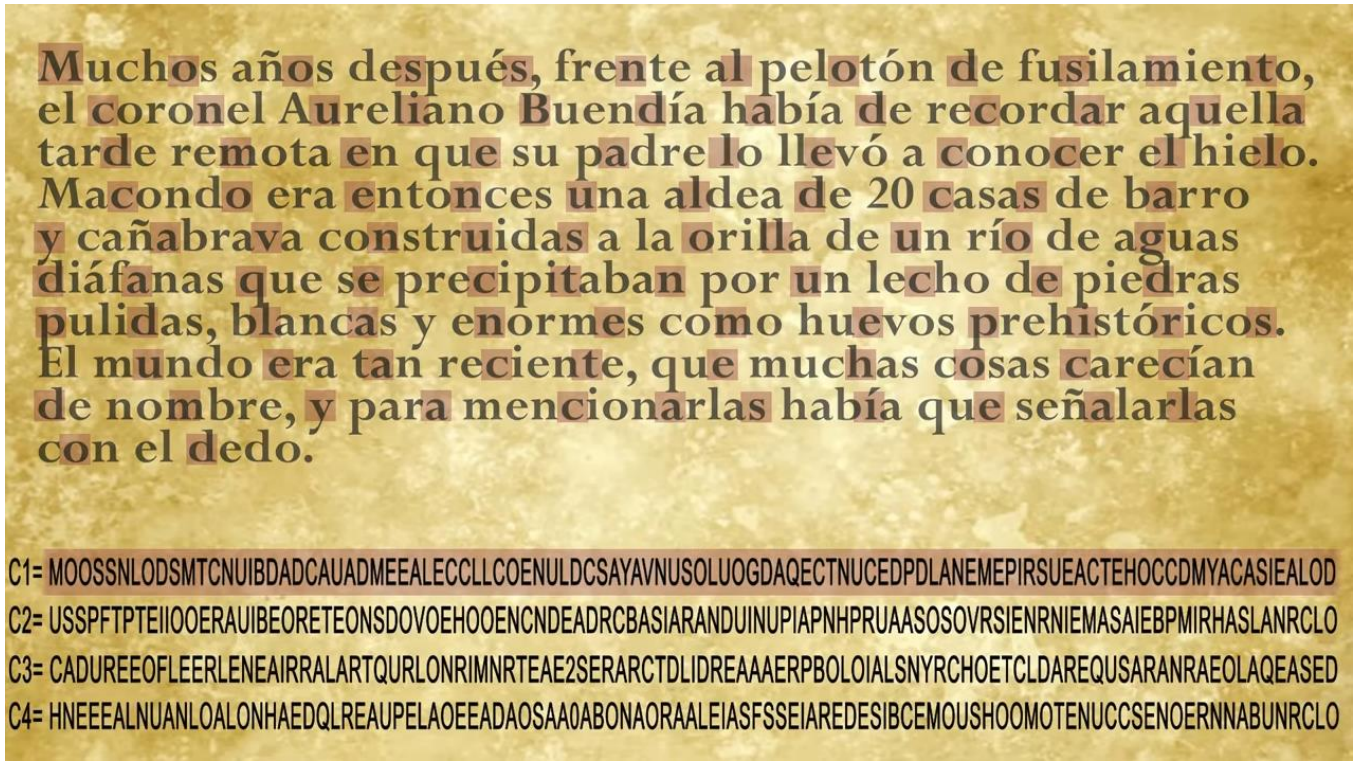
Clave utilizada: SECRETOS

Ahora que todo acabó y que el tiempo te ha vencido,
y tu amigo te dejó, dices que cuentas conmigo.
Como tienes el valor, yo que siempre me he dolido
de recordar lo que fue y lo que pudo haber sido.
Por la calle del olvido vagan tu sombra y la mía,
cada una en una acera por las cosas de la vida.
Por la calle del olvido donde nunca brilla el día,
condenados a una noche tan oscura como fría.

No sabes lo que luché para no soñar contigo
y no quieres entender que por fin lo he conseguido.
Yo estaba dispuesto a todo para tenerte conmigo
hasta hubiera trabajado, y te fuiste con mi amigo.
Por la calle del olvido vagan tu sombra y la mía,
cada una en una acera por las cosas de la vida.
Por la calle del olvido donde nunca brilla el día,
condenados a una noche tan oscura como fría.
La calle del olvido. Los Secretos.

https://www.youtube.com/watch?v=O0_b4bJU-EQ

Más información en píldoras Thoth



https://www.youtube.com/watch?v=A7p2ydEPg1k

Conclusiones de la lección 6.6

- La cifra de Vigenère inventada en 1586, se criptoanaliza 277 años después en 1863 mediante el método de Kasiski
- No obstante en 1854, nueve años antes, Charles Babbage logra romper la cifra Autoclave, algo más complicada que Vigenère
- El método de Kasiski busca primero encontrar la longitud L de la clave
- Conocida esa supuesta longitud L, se rompe el criptograma de L en L letras
- Se sabe que con cada posición de esa clave la cifra se vuelve monoalfabética
- Y lo novedoso es que la redundancia del lenguaje se sigue manifestando si un texto se lee “a saltos”, de L en L letras
- Aplicamos ahora en cada uno de los subcriptogramas la regla AEOS, donde la posición relativa de la A marcará la letra de la clave

Lectura recomendada

- Introducción a la seguridad informática y criptografía clásica. Lección 9: Algoritmos de cifra por sustitución polialfabética, apartados de criptoanálisis, MOOC Crypt4you, Jorge Ramió, 2017
  - https://www.criptored.es/crypt4you/temas/criptografiaclasica/leccion9.html
- Guion píldora formativa Thoth nº 20, ¿Cómo se ataca por Kasiski la cifra de Vigenère?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora020.pdf
- Cifrado Autoclave y criptoanálisis, Criptografía clásica, Jorge Ramió, 2006
  - https://www.criptored.es/descarga/CriptoClasicapdf.zip
- The index of coincidence, William F. Friedman, 1922
  - https://www.nsa.gov/Portals/70/documents/news-features/declassified-documents/friedman-documents/publications/FOLDER_233/41761039080018.pdf

Class4crypt c4c6.7

Módulo 6. Algoritmos de criptografía clásica

Lección 6.7. Cifrado digrámico de Playfair

6.7.1. Monogramas y digramas

6.7.2. Formando bloques de letras

6.7.3. Charles Wheatstone y Lyon Playfair

6.7.4. Cifrado digrámico por cuadrado de Playfair

6.7.5. Descifrado digrámico por cuadrado de Playfair

6.7.6. Criptoanálisis a la cifra de Playfair

Class4crypt c4c6.7 Cifrado digrámico de Playfair
https://www.youtube.com/watch?v=bagPJ-SMO1c

Monogramas y digramas

- Rarezas del lenguaje
  - La RAE no recoge monograma, digrama, … poligrama
    - *Grama: Significa 'escrito' o 'gráfico'. Cablegrama, cardiograma, epigrama*
  - En inglés a esta cifra se le llama polygraphic substitution
    - Buscamos entonces en la RAE por poligráfica, poligrafía
      - *Arte de escribir por diferentes modos secretos o extraordinarios, de suerte que lo escrito no sea inteligible sino para quien pueda descifrarlo*
      - *Arte de descifrar los escritos poligráficos*
      - ¿Qué es escribir de modo “extraordinario”?
  - Deberíamos asociar “grama” más a una letra que a un escrito

Formando bloques de letras o poligramas

- La cifra por sustitución monoalfabética monográmica, letra a letra, se mostraba muy débil porque las estadísticas de las letras del alfabeto se manifestaban de forma muy clara en el criptograma
- Incluso la cifra por sustitución polialfabética monográmica, mucho más segura que la anterior, al final también sucumbía ante un ataque de tipo estadístico reducido a una cifra monoalfabética
- A mediados del siglo XIX, el cifrado de Wheatstone, más conocido como cifrado de Playfair, propone cifrar bloques de dos letras
- Cifrando digramas se dificulta el ataque, porque las estadísticas de los 729 digramas están más dispersas que la de los monogramas

Charles Wheatstone y Lyon Playfair

- Playfair cifra digramas con una matriz de 5x5 letras, desde donde se va formando el criptograma
- Inventado en 1854 por el inventor británico Charles Wheatstone, fue promovido para su uso militar por el científico también británico Lyon Playfair
- Entre otros escenarios, las fuerzas británicas lo usan en la primera y segunda guerras mundiales
- Su seguridad radica en que cifra bloques de dos letras o digramas, y en este caso el criptoanálisis por el estudio de la frecuencia de digramas tiene una mayor dificultad

Charles Wheatstone

Lyon Playfair

El cuadrado de Playfair módulo 27

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

- Se ponen las 27 letras del alfabeto en las celdas del cuadrado de Playfair
- Como hay solamente 25 celdas, dos letras de semejante apariencia, como pueden ser la I con la J y la N con la Ñ, podrían ir juntas en la misma celda

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

- O bien se sustituyen la Ñ por la N y la J por la I en el texto en claro, que es como trabaja el software Criptoclásicos v2.1
- Del texto en claro se formarán bloques de dos letras M_1M_2 para crear el criptograma C_1C_2

Cifrando con cuadrado de Playfair mod 27

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

- Para darle seguridad usaremos una clave
- Si M_1M_2 están en la misma fila, C_1C_2 serán las letras inmediatamente a la derecha de M_1 y de M_2 en módulo 5 en dicha fila
- Si M_1M_2 están en la misma columna, C_1C_2 serán las letras inmediatamente hacia abajo de M_1 y de M_2 en módulo 5 en dicha columna
- Si M_1M_2 están en fila y columna diferentes, C_1C_2 serán las letras de la diagonal opuesta, partiendo siempre desde la fila del primer elemento M_1
- Si el último bloque tiene una sola letra, se usará una letra como relleno y si un bloque M_1M_2 tiene dos letras iguales, para romper ese cifrado repetido, se incluirá entre ellas la misma letra que se use como relleno
- Para el descifrado se realizará la operación inversa

Ejemplo de cifrado de Playfair mod 27

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

- Clave: SOLAR, Texto: RELLENOS AQUI
- Bloques de digramas M_1M_2
- RE LL EN OS AQ UI (romperemos la doble L)
- M_1M_2 = RE LX LE NO SA QU IX (otro relleno X)
- C = AF DL AD PS OR TN QL

- RE se cifra como AF, la diagonal opuesta desde la fila de M_1 = R
- LX se cifra como DL, letras de abajo en la misma columna módulo 5
- LE se cifra como AD, la diagonal opuesta desde la fila de M_1 = L
- NO se cifra como PS, la diagonal opuesta desde la fila de M_1 = N
- SA se cifra como OR, letras a la derecha en la misma fila módulo 5 … etc.

Ejemplo de descifrado de Playfair mod 27

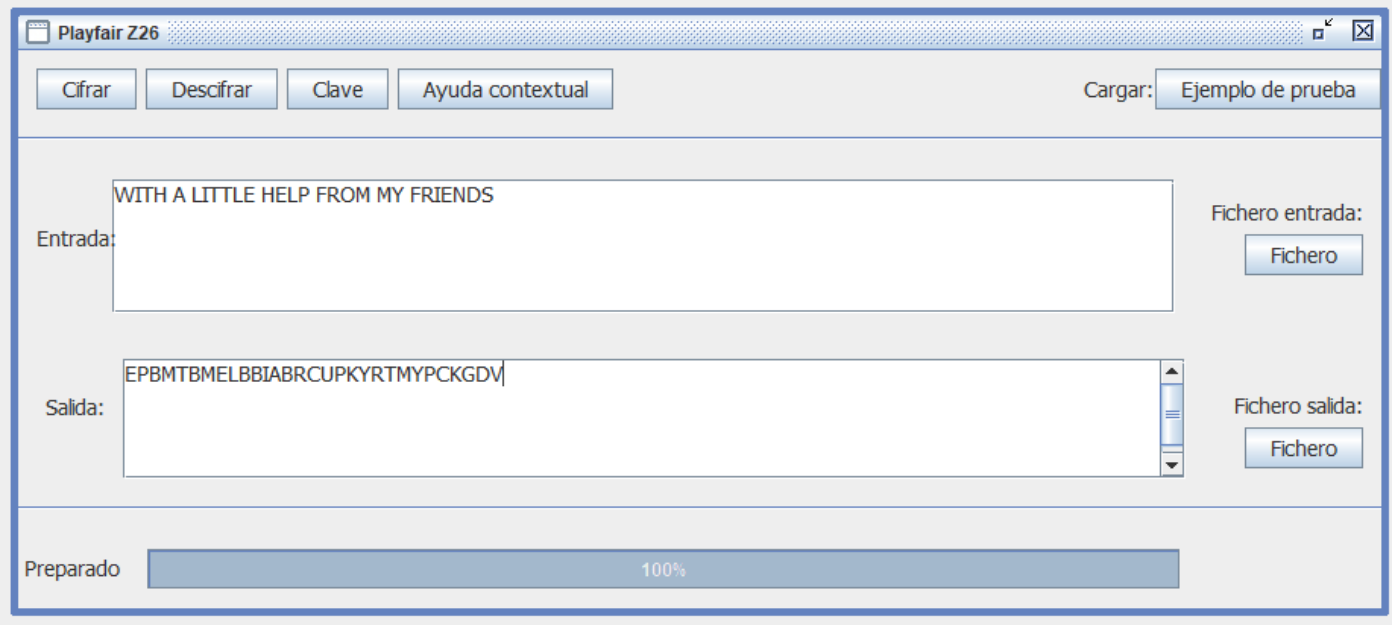
S	O	L	A	R
B	C	D	E	F
G	H	I	K	M
N	P	Q	T	U
V	W	X	Y	Z

- Clave: SOLAR, C: AF DL AD PS OR TN QL
- Para descifrar, si C_1C_2 están en la misma fila, se toman las letras de la izquierda mod 5; si están en la misma columna, se toman las letras de arriba mod 5; y si están en fila y columna diferentes, se toman las letras de la diagonal opuesta siempre partiendo de C_1
- AF se descifra como RE, la diagonal opuesta desde la fila de C_1 = A
- DL se descifra como LX, letras de arriba en la misma columna módulo 5
- AD se descifra como LE, la diagonal opuesta desde la fila de C_1 = A
- PS se descifra como NO, la diagonal opuesta desde la fila de C_1 = P
- OR se descifra como SA, letras a la izquierda en la misma fila módulo 5 … etc.

Ejemplo con Criptoclásicos v2.1 mod 26

- Clave: BEATLES
- Texto: WITH A LITTLE HELP FROM MY FRIENDS
- Digramas: WI TH AL IT TL EH EL PF RO MX MY FR IE ND SX

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



https://www.criptored.es/software/sw_m001c.htm

C = EP BM TB ME LB BI AB RC UP KY RT MY PC KG DV

Agradecimientos a:
Juan Contreras Rubio, por sus actualizaciones
https://www.linkedin.com/in/juancr87/

Criptoanálisis a la cifra de Playfair

- Se buscan digramas con alta frecuencia en el criptograma. Estos digramas deberían proceder de digramas frecuentes en el lenguaje
- También va a ayudar en el criptoanálisis el hecho de que en el lenguaje hay digramas frecuentes cuya lectura inversa también es frecuente, como por ejemplo AR y RA (EN TR EN <u>AR</u> <u>RA</u> PI DO), SE y ES (<u>SE</u> <u>ES</u> PE RA UN PO CO), etc., que se cifrarán también de forma inversa
- A partir de ello habrá que ir formando un posible cuadrado de Playfair
- Existen otros métodos diferentes de criptoanálisis. Encontrará muchos documentos e informes en Internet, no lo veremos aquí
- Si se tiene el texto en claro y el texto cifrado, no es muy difícil encontrar la clave (ver la bibliografía)

Conclusiones de la Lección 6.7

- Los cifrados digrámicos complican el criptoanálisis por la mayor dificultad de aplicar estadísticas del lenguaje a los digramas en vez de los monogramas
- El cifrado de Playfair (1854) fue el primero en cifrar bloques de dos letras y se usó en las dos guerras mundiales por parte de las fuerzas británicas
- En el cuadrado de Playfair, de 5 filas por 5 columnas, se escriben 25 letras del alfabeto con una clave, procediendo a la cifra de bloques M_1M_2 como se indica
  - Si M_1M_2 están en la misma fila, C_1C_2 serán las letras a la derecha de la fila mod 5
  - Si M_1M_2 están en la misma columna, C_1C_2 serán las letras de abajo de la columna mod 5
  - Si M_1M_2 están en fila y columnas distintas, C_1C_2 serán las letras de la diagonal opuesta partiendo siempre desde la fila de la letra M_1
  - Para el descifrado, se usarán las operaciones inversas sobre C_1C_2: a la izquierda de la fila, hacia arriba de la columna y la diagonal opuesta ahora desde la fila de la letra C_1

Lectura recomendada

- Libro electrónico de seguridad informática y criptografía, versión 4.1, Jorge Ramió, 2006, archivo Criptografía clásica (Word y PDF), 105 páginas
  - https://www.criptored.es/guiateoria/gt_m001a.htm
- Playfair cipher, Wikipedia
  - https://en.wikipedia.org/wiki/Playfair_cipher
- Cryptanalysis of the Playfair cipher, Practical cryptography
  - http://www.practicalcryptography.com/cryptanalysis/stochastic-searching/cryptanalysis-playfair/
- How to find the keyword of the Playfair cipher, given the plaintext and the ciphertext?, Cryptography
  - https://crypto.stackexchange.com/questions/35722/how-to-find-the-keyword-of-the-playfair-cipher-given-the-plaintext-and-the-ciph

Class4crypt c4c6.8

Módulo 6. Algoritmos de criptografía clásica

Lección 6.8. Cifrado poligrámico con matrices de Hill

6.8.1. Cifrado poligrámico de Hill

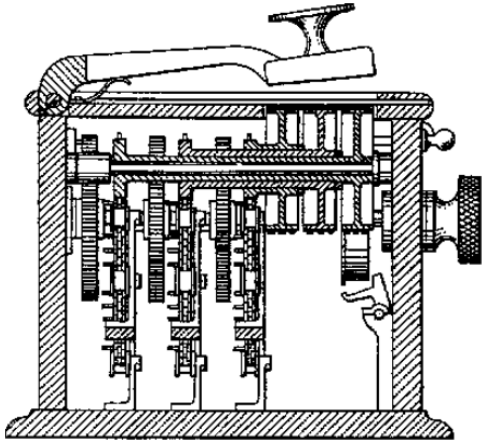
6.8.2. Fortalezas y debilidades de la cifra de Hill

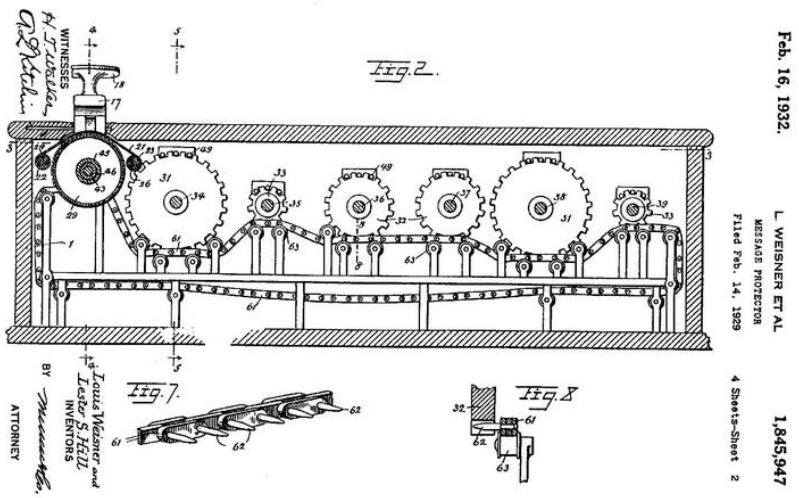
6.8.3. Uso de bloques en la cifra moderna

Class4crypt c4c6.8 Cifrado poligrámico con matrices de Hill
https://www.youtube.com/watch?v=k74FLglT1Ns

Lester Hill

- Matemático norteamericano que propone en el año 1929 el uso de matrices para la cifra de dos o más letras del texto en claro





- Hill patenta su invento con claves de 6 letras, pero como dicha clave debe ser fija para cada máquina, no puede competir con otras máquinas de la época como Enigma, Lorenz o Purple

Propuesta original de Hill (1/3)

- Toma las 26 letras del alfabeto inglés y las codifica de una manera particular

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

mod 26

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
05	23	02	20	10	15	08	04	18	25	00	16	13	07	03	01	19	06	12	24	21	17	14	22	11	09

- Plantea esta matriz de 4x4, siendo x_i el texto en claro e y_i el criptograma

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 8 & 6 & 9 & 5 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \bmod 26$$

Matriz clave K mod 26

- $y_1 = 08 * x_1 + 06 * x_2 + 09 * x_3 + 05 * x_4 \bmod 26$
- $y_2 = 06 * x_1 + 09 * x_2 + 05 * x_3 + 10 * x_4 \bmod 26$
- $y_3 = 05 * x_1 + 08 * x_2 + 04 * x_3 + 09 * x_4 \bmod 26$
- $y_4 = 10 * x_1 + 06 * x_2 + 11 * x_3 + 04 * x_4 \bmod 26$

Propuesta original de Hill (2/3)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
05	23	02	20	10	15	08	04	18	25	00	16	13	07	03	01	19	06	12	24	21	17	14	22	11	09

mod 26

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 8 & 6 & 9 & 5 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \text{ mod } 26$$

Matriz clave K mod 26

- M: DELAY OPERATIONS
- M: DELA YOPE RATI ONSU (relleno U)

  Bloque 1: 20 10 16 05 Bloque 2: 11 03 01 10

  Bloque 3: 06 05 24 18 Bloque 4: 03 07 12 21

- y_1 = 08 * 20 + 06 * 10 + 09 * 16 + 05 * 05 mod 26 = 160 + 60 + 144 + 25 = 389 mod 26 = 25 = J
- y_2 = 06 * 20 + 09 * 10 + 05 * 16 + 10 * 05 mod 26 = 120 + 90 + 80 + 50 = 340 mod 26 = 02 = C
- y_3 = 05 * 20 + 08 * 10 + 04 * 16 + 09 * 05 mod 26 = 100 + 80 + 64 + 45 = 289 mod 26 = 03 = O
- y_4 = 10 * 20 + 06 * 10 + 11 * 16 + 04 * 05 mod 26 = 200 + 60 + 176 + 20 = 456 mod 26 = 14 = W
- C = JCOW ZLVB DVLE QMXC

¿Cómo desciframos ahora C?

Propuesta original de Hill (3/3)

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
05	23	02	20	10	15	08	04	18	25	00	16	13	07	03	01	19	06	12	24	21	17	14	22	11	09

mod 26

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 23 & 20 & 5 & 1 \\ 2 & 11 & 18 & 1 \\ 2 & 20 & 6 & 25 \\ 25 & 2 & 22 & 25 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \text{ mod } 26$$

Matriz clave inversa K^{-1} mod 26

- C: JCOW ZLVB DVLE QMXC

  Bloque 1: 25 02 03 14 Bloque 2: 09 16 17 23

  Bloque 3: 20 17 16 10 Bloque 4: 19 13 22 02

- x_1 = 23 * 25 + 20 * 02 + 05 * 03 + 01 * 14 mod 26 = 575 + 40 + 15 + 14 = 644 mod 26 = 20 = D
- x_2 = 02 * 25 + 11 * 02 + 18 * 03 + 01 * 14 mod 26 = 50 + 22 + 54 + 14 = 140 mod 26 = 10 = E
- x_3 = 02 * 25 + 20 * 02 + 06 * 03 + 25 * 14 mod 26 = 50 + 40 + 18 + 350 = 458 mod 26 = 16 = L
- x_4 = 25 * 25 + 02 * 02 + 22 * 03 + 25 * 14 mod 26 = 625 + 4 + 66 + 350 = 1.045 mod 26 = 05 = A
- M = DELA YOPE RATI ONSU

Cifrado de Hill genérico mod 27

Cifrado del texto M
(C) = (K) x (M) mod 27

Descifrado del criptograma C
(M) = (K^{-1}) x (C) mod 27

$$\begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \dots \\ c_N \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} & k_{13} & & k_{1N} \\ k_{21} & k_{22} & k_{23} & & k_{2N} \\ k_{31} & k_{32} & k_{33} & & k_{3N} \\ \dots & \dots & \dots & & \dots \\ k_{N1} & k_{N2} & k_{N3} & & k_{NN} \end{pmatrix} \begin{pmatrix} m_1 \\ m_2 \\ m_3 \\ \dots \\ m_N \end{pmatrix} \bmod 27$$

- La matriz clave K debe tener inversa K^{-1} en el módulo de cifra n
- Como $K^{-1} = T_{ADJ(K)}/|K|$ mod n, en donde ADJ(K) es la matriz adjunta, T es la transpuesta y |K| el determinante. Este último valor |K| no podrá ser cero ni tener factores en común con n (condición de inversos)
- Si el texto en claro no es múltiplo del bloque de N letras, el último bloque se rellenará con una letra predeterminada de poco uso

Matrices simbólicas mod 27

- Si la matriz clave K tiene un orden grande, es poco práctico enviar los NxN elementos de la misma
- Aquí es recomendable usar lo que se denomina matriz simbólica, en donde los elementos k_{ij} de la matriz K se sustituyen por una palabra o un texto

$$\begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \text{ mod } 27$$

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

- Matriz clave de 3x3 simbólica: CANCIONES
  - Fila 1 (C A N), Fila 2 (C I O), Fila 3 (N E S)
  - Determinante = 16 (matriz clave válida)
- Pero CAMIONERO no nos servirá como matriz clave en mod 27 porque su determinante es 18

$$\begin{pmatrix} 02 & 00 & 13 \\ 02 & 08 & 15 \\ 13 & 04 & 19 \end{pmatrix} \text{ mod } 27$$

Cifrado de Hill en mod 27

- Texto en claro M = ESCRIBO LAS CANCIONES QUE HACEN CANTAR AL MUNDO ENTERO (I Write The Songs - Barry Manilow - 1975)
- Matriz clave simbólica 3x3: CANCIONES

$$[K]\begin{pmatrix} 02 & 00 & 13 \\ 02 & 08 & 15 \\ 13 & 04 & 19 \end{pmatrix} \bmod 27$$

C = HBE VHO DKW LAM VAF BFZ ÑDM
IKZ LOS LKD ENI UYC ACE KÑZ ERX ROR

C_1 = 02*E + 00*S + 13*C = 2*4 + 0*19 + 13*2 = 34 mod 27 = 7 = H
C_2 = 02*E + 08*S + 15*C = 2*4 + 8*19 + 15*2 = 190 mod 27 = 1 = B
C_3 = 13*E + 04*S + 19*C = 13*4 + 4*19 + 19*2 = 166 mod 27 = 4 = E

Descifrado de Hill en mod 27

- Criptograma C = HBE VHO DKW LAM VAF BFZ ÑDM IKZ LOS LKD ENI UYC ACE KÑZ ERX ROR
- Matriz clave simbólica 3x3: CANCIONES

$$\left(K^{-1}\right)\begin{pmatrix} 26 & 10 & 07 \\ 25 & 07 & 20 \\ 21 & 13 & 01 \end{pmatrix} \bmod 27$$

M = ESC RIB OLA SCA NCI ONE SQU EHA CEN CAN TAR ALM UND OEN TER OXX

$M_1 = 26*H + 10*B + 07*E = 26*7 + 10*1 + 7*4 = 220 \bmod 27 = 4 = E$

$M_2 = 25*H + 07*B + 20*E = 25*7 + 7*1 + 20*4 = 262 \bmod 27 = 19 = S$

$M_3 = 21*H + 13*B + 01*E = 21*7 + 13*1 + 1*4 = 164 \bmod 27 = 2 = C$

Ejemplo práctico con Criptoclásicos v2.1

- Con la matriz de cifra K de 4x4, se pide cifrar mod 27 el texto M = ERASE UNA VEZ EN EL LEJANO OESTE
- C = IKÑY EUHÑ PXBM QLUH TPLK YUDU UBDN
- Descifrar el criptograma anterior
- M = ERAS EUNA VEZE NELL EJAN OOES TEZZ

$$\begin{pmatrix} 01 & 02 & 03 & 04 \\ 05 & 06 & 07 & 08 \\ 09 & 21 & 22 & 23 \\ 24 & 25 & 26 & 01 \end{pmatrix}$$

$$\begin{pmatrix} 100 & 101 & 102 \\ 103 & 104 & 103 \\ 102 & 101 & 100 \end{pmatrix}$$

- Descifrar C en mod 191 con la matriz K de 3x3
- C = r]0· C§¤K jájW]Ý9e GYç§ å·ªÏ Q1WÁ hyF¡ 9ÕiË <ó*Ñ a¹µ" í@Î@ pù¢u %}¯\ «ZJd Áhyð ®ÜIï MFôf ÝH½J ÁP·ª Ï°cª 30/ð ®ÜKò 12áý ¥du3 ô32á ý»¿± 1oMÕ ÂÝñ± Û¬vD °Çý# qÎ)Á +2´" ;xüå Öãðø fË°Ñ ÇGÃG &e+/ å~ë?

Fortalezas y debilidades de Hill (1/2)

- De todas las matrices de clave K posibles, tenemos que eliminar aquellas en las que su determinante:
  a) Sea igual a cero (matrices singulares)
  b) Tenga factores en común con el módulo n de cifra
- La pregunta es, ¿qué sucedería con la matriz clave K si usamos un módulo de cifra cuyo número sea un primo p?
- Lo que va suceder es que, excepto aquellas matrices en las que el determinante sea igual cero, que en realidad serán muy pocas, todas las demás matrices serán válidas, porque cada uno de los p-1 restos de un primo p tendrán su correspondiente inverso multiplicativo

Fortalezas y debilidades de Hill (2/2)

- Si n = p, el número de matrices válidas $K_{Válidas}$ tiende al máximo p^{N^2}
  - Si hacemos una cifra con bloques de 8 letras y usamos como módulo p = 29 (por ejemplo las mayúsculas y los signos + y -), el número de matrices válidas $K_{Válidas} = 29^{64} = 3{,}9 \times 10^{93} \approx 2^{310}$
  - Este espacio de claves será muy superior (2^{54} veces mayor) a la del algoritmo AES 256, que tiene 2^{256} claves posibles
  - Si p = 191 (un subconjunto imprimible del código ASCII, definido en Criptoclásicos), con un bloque de 8 caracteres, el número de matrices válidas se eleva a $K_{Válidas} = 191^{64} = 9{,}7 \times 10^{145} \approx 2^{485}$
  - Pero este sistema será muy vulnerable ante un criptoanálisis por texto en claro conocido, mediante el método de Gauss-Jordan

Cifra poligrámica en criptografía moderna

- Igual como sucedía con la cifra monográmica clásica, letra a letra, que en la criptografía moderna se cifraba bit a bit (e.g. A5) o byte a byte (e.g. RC4) en la denominada cifra simétrica en flujo, o stream ciphers, va a suceder lo mismo con la cifra poligrámica clásica y la moderna
- En la criptografía moderna usaremos bloques de texto en claro para la cifra simétrica en bloque, agrupando el texto en bytes
  - El algoritmo DES cifra bloques de texto de 64 bits (8 bytes)
  - El algoritmo IDEA cifra bloques de texto de 64 bits (8 bytes)
  - El algoritmo AES cifra bloques de texto de 128 bits (16 bytes)
  - El porqué de esos tamaños en los bloques de cifra se explicará en la clase de introducción a la cifra simétrica en bloque

Más información en píldoras Thoth

https://www.youtube.com/watch?v=3X29bcufrOM

Conclusiones de la Lección 6.8

- Los cifrados poligrámicos complican el criptoanálisis por la mayor dificultad en aplicar estadísticas del lenguaje a digramas, trigramas, etc.
- El cifrado de Hill (1929) permite cifrar bloques de texto en claro mayores que los digramas de Playfair, haciendo uso de matrices
- Para cifrar, la matriz clave K se multiplica por la matriz de texto en claro reduciendo mod n. Para descifrar, la matriz inversa de la clave K^{-1} se multiplica por la matriz del criptograma reduciendo mod n
- Para que exista la matriz inversa de descifrado K^{-1}, el determinante de la matriz clave K no debe ser 0 ni tener factores con el módulo n
- Aunque el espacio de claves puede ser inmenso, el cifrado de Hill no soportará un ataque con texto en claro conocido (próxima clase)

Lectura recomendada

- Guion píldora formativa Thoth nº 21, ¿Qué es la cifra por matrices de Hill?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora021.pdf
- Libro electrónico de seguridad informática y criptografía, versión 4.1, Jorge Ramió, 2006, archivo Criptografía clásica (Word y PDF), 105 páginas
  - https://www.criptored.es/guiateoria/gt_m001a.htm
- Hill cipher, Wikipedia
  - https://en.wikipedia.org/wiki/Hill_cipher

Class4crypt c4c6.9

Módulo 6. Algoritmos de criptografía clásica

Lección 6.9. Criptoanálisis a la cifra de Hill por Gauss-Jordan

6.9.1. Buscando los vectores unitarios

6.9.2. Las figuras de Gauss y Jordan

6.9.3. Criptoanálisis de Gauss-Jordan

6.9.4. Ejemplo práctico mod 27 y mod 191

6.9.5. Usos de matrices en la criptografía moderna

Class4crypt c4c6.9 Criptoanálisis a la cifra de Hill por Gauss-Jordan
https://www.youtube.com/watch?v=D2gyofZxAl0

Los vectores unitarios en Hill

- Se conoce como vector unitario a la fila de una matriz que contiene un uno en la columna que le corresponde en una matriz identidad y los demás valores son ceros
- Así, la fila {B A A A … A A} sería el vector unitario {1 0 0 0 … 0 0}
- Si un bloque de texto se cifra con una matriz identidad, es decir con los vectores unitarios, se obtiene el mismo texto en claro
- Por ejemplo si el trigrama HOL se cifra con una matriz identidad

$$\begin{pmatrix} B & A & A \\ A & B & A \\ A & A & B \end{pmatrix} \begin{pmatrix} H \\ O \\ L \end{pmatrix} \bmod 27 \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 7 \\ 15 \\ 11 \end{pmatrix} \bmod 27$$

$C_1 = 1*7 + 0*15 + 0*11 \bmod 27 = 7 = H$
$C_2 = 0*7 + 1*15 + 0*11 \bmod 27 = 15 = O$
$C_3 = 0*7 + 0*15 + 1*11 \bmod 27 = 11 = L$

Buscando vectores unitarios en M (1/2)

- Si la matriz clave K es de 2x2 no es tan difícil encontrar los vectores unitarios AB (0 1) y BA (1 0) en el texto en claro
  - M = NOS HABLABA EN VOZ ALTA M = NO SH AB LA BA EN VO ZA LT AZ
  - Cifraremos M con la clave (K) mod 27 con $k_{11} = 7$, $k_{12} = 5$, $k_{21} = 12$, $k_{22} = 19$
  - C = EJ GK FS WX HM MY NJ TO OZ VI
  - Por lo tanto si sabemos que AB se cifra como FS, podremos conocer k_{12} y k_{22}

$$\begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \bmod 27 = \begin{pmatrix} F \\ S \end{pmatrix}$$

- Si $k_{11}*A + k_{12}*B \bmod 27 = F$
  - $k_{11}*0 + k_{12}*1 \bmod 27 = F = 5$
  - $k_{12} = 5$
- Si $k_{21}*A + k_{22}*B \bmod 27 = S$
  - $k_{21}*0 + k_{22}*1 \bmod 27 = S = 19$
  - $k_{22} = 19$

Buscando vectores unitarios en M (2/2)

- M = NO SH AB LA BA EN VO ZA LT AZ
- C = EJ GK FS WX HM MY NJ TO OZ VI
  - Y además sabemos que BA se cifra como HM, podremos conocer k_{11} y k_{21}

$$\begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} B \\ A \end{pmatrix} \bmod 27 = \begin{pmatrix} H \\ M \end{pmatrix}$$

- Si $k_{11}*B + k_{12}*A \bmod 27 = H$
  - $k_{11}*1 + k_{12}*0 \bmod 27 = H = 7$
  - $k_{11} = 7$
- Si $k_{21}*B + k_{22}*A \bmod 27 = M$
  - $k_{21}*1 + k_{22}*0 \bmod 27 = M = 12$
  - $k_{21} = 12$

$$\begin{pmatrix} 7 & 5 \\ 12 & 19 \end{pmatrix}$$

Era la matriz clave usada en la cifra

Si la matriz clave (K) es de 3x3 o mayor, es bastante más difícil encontrar estos vectores unitarios en el texto en claro: BAA, ABA, AAB

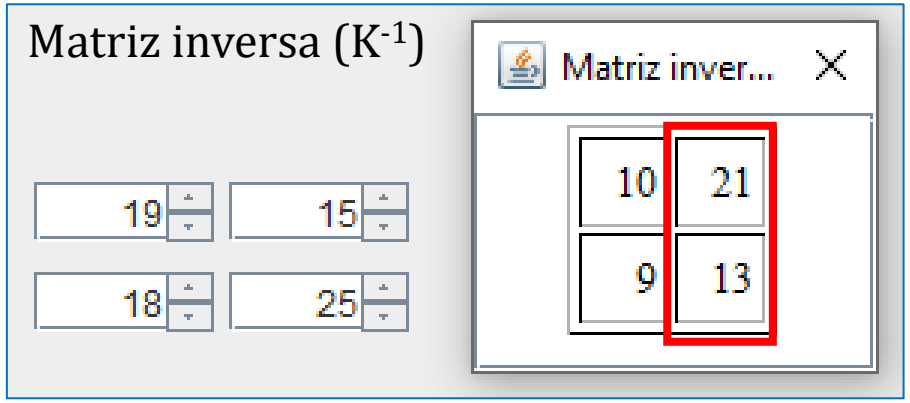
Buscando vectores unitarios en C

- Otra opción sería buscar estos vectores unitarios en el criptograma
  - C = KD NÑ GB MX SK ÑW OR KH AB TW AR ZG PK KK WF DL KF XO RA MA
  - M = BU SC AN DO VE CT OR ES UN IT AR IO SE NE LC RI PT OG RA MA

$$\begin{pmatrix} k^{-1}_{11} & k^{-1}_{12} \\ k^{-1}_{21} & k^{-1}_{22} \end{pmatrix} \begin{pmatrix} A \\ B \end{pmatrix} \bmod 27 = \begin{pmatrix} U \\ N \end{pmatrix}$$

Matriz (K) usada en esta cifra $\begin{pmatrix} 19 & 15 \\ 18 & 25 \end{pmatrix}$

- Si $k^{-1}_{11}*A + k^{-1}_{12}*B \bmod 27 = U$
  - $k^{-1}_{11}*0 + k^{-1}_{12}*1 \bmod 27 = U = 21$
  - $k^{-1}_{12} = 21$
- Si $k^{-1}_{21}*A + k^{-1}_{22}*B \bmod 27 = N$
  - $k^{-1}_{21}*0 + k^{-1}_{22}*1 \bmod 27 = N = 13$
  - $k^{-1}_{22} = 13$



La clave está en los vectores unitarios

- Con los vectores unitarios en el texto en claro o en el criptograma, seremos capaces de conocer los elementos de la clave (K)
- Debemos conocer cómo se ha cifrado cada bloque de texto, es decir, se trata de un ataque o criptoanálisis por texto en claro conocido
- Pero no es tan fácil encontrar estos vectores unitarios en el texto en claro ni en el criptograma, y mucho menos si el orden de la matriz clave (K) es de 3x3 o superior
- ¿Qué podríamos hacer entonces?
- Forzar a que aparezcan dichos vectores usando el método de Gauss-Jordan
- Se puede hacer porque cada bloque de texto en claro (digrama, trigrama, etc.) se cifra de forma independiente y permite que se hagan operaciones modulares sobre ese bloque de texto en claro y su criptograma asociado

Johann Karl Gauss y Wilhelm Jordan

- Johann Karl Friedrich Gauss (1777-1855), destacado matemático alemán que contribuyó al desarrollo de la teoría de números: en 1798 con 21 años escribe *Disquisitiones arithmeticae* que se publica en 1801

- Wilhelm Jordan (1842–1899), geodesista alemán, que publica en *Handbuch der Vermessungskunde* (1873) la técnica conocida como *Eliminación de Gauss-Jordan* en honor a ellos, que permite encontrar matrices inversas
- Aplicando este método, se rompe la clave del cifrado de Hill

Criptoanálisis por método de Gauss-Jordan

- El método consiste en escribir una matriz 2N-grámica con los elementos del texto en claro y los elementos del criptograma
- En esta matriz realizaremos operaciones lineales (multiplicar filas por un número para buscar el valor 1 y restar filas entre sí para buscar los demás valores 0), es decir los vectores unitarios
- Una vez se logra en una de esas dos matrices una matriz identidad, de la otra matriz se podrá deducir la matriz de clave (K)
- Si la matriz 2N-grámica es [Texto en claro] | [Texto cifrado], al lograr una matriz identidad en [Texto en claro], la matriz resultante de [Texto cifrado] será la traspuesta de la matriz clave (K), es decir, simplemente debemos cambiar filas por columnas

Ejemplo del método de Gauss-Jordan

- M = LAV IDA ESC OMO UNA CAJ ADE BOM BON ESN UNC AZZ …
- C = NLT GMW IQD OXD OBI GJT PFO OMV BIN PAM ÑTQ ZGK …

A	B	C	D	E	F	G	H	I	J	K	L	M	N	Ñ	O	P	Q	R	S	T	U	V	W	X	Y	Z
00	01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26

$$\left(\begin{array}{ccc|ccc} L & A & V & N & L & T \\ I & D & A & G & M & W \\ E & S & C & I & Q & D \\ O & M & O & O & X & D \\ U & N & A & O & B & I \\ C & A & J & G & J & T \\ .. & .. & .. & .. & .. & .. \end{array}\right) = \left(\begin{array}{ccc|ccc} 11 & 0 & 22 & 13 & 11 & 20 \\ 8 & 3 & 0 & 6 & 12 & 23 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{ mod } 27$$

- En la primera fila vamos a dejar el primer elemento (ahora un 11) como un 1, multiplicando toda esa fila por el inv (11, 27) = 5
- Si esto no puede hacerse porque no exista el inverso del número en mod 27, movemos filas completas

Forzando el 1 del vector unitario (1 0 0)

$$\left(\begin{array}{ccc|ccc} 11 & 0 & 22 & 13 & 11 & 20 \\ 8 & 3 & 0 & 6 & 12 & 23 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{mod } 27$$

Multiplicamos la fila 1 por el inverso de 11 en módulo 27, que es 5

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 8 & 3 & 0 & 6 & 12 & 23 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{mod } 27$$

- 5*11 mod 27 = 1 5* 0 mod 27 = 0 5*22 mod 27 = 2
- 5*13 mod 27 = 11 5*11 mod 27 = 1 5*20 mod 27 = 19

- Ahora debemos dejar toda la columna 1, desde la fila 2 hacia abajo, con ceros

Poniendo 0 en la 2ª fila de la 1ª columna

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 8 & 3 & 0 & 6 & 12 & 23 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{mod } 27$$

Hacemos:
2ª fila =
2ª fila – 8 * 1ª fila

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 3 & 11 & 26 & 4 & 6 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{mod } 27$$

- 8 – 8*1 mod 27 = 0 3 – 8*0 mod 27 = 3 0 – 8*2 mod 27 = 11
- 6 – 8*11 mod 27 = 26 12 – 8 *1 mod 27 = 4 23 – 8*19 mod 27 = 6

- Repetimos ahora esta operación de resta con la tercera fila

Poniendo 0 en la 3ª fila de la 1ª columna

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 3 & 11 & 26 & 4 & 6 \\ 4 & 19 & 2 & 8 & 17 & 3 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{ mod } 27$$

Hacemos
3ª fila =
3ª fila – 4 * 1ª fila

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 3 & 11 & 26 & 4 & 6 \\ 0 & 19 & 21 & 18 & 13 & 8 \\ 15 & 12 & 15 & 15 & 24 & 3 \\ 21 & 13 & 0 & 15 & 1 & 8 \\ 2 & 0 & 9 & 6 & 9 & 20 \\ .. & .. & .. & .. & .. & .. \end{array}\right) \text{ mod } 27$$

- 4 – 4*1 mod 27 = 0 19 – 4*0 mod 27 = 19 2 – 4*2 mod 27 = 21
- 8 – 4*11 mod 27 = 18 17 – 4 *1 mod 27 = 13 3 – 4*19 mod 27 = 8

- Pasamos la 3ªfila como 2ª porque, a diferencia del 3, el 19 sí tiene inverso en módulo 27 y eso nos sirve para encontrar para vector unitario (0 1 0)

Forzando el 1 del vector unitario (0 1 0)

- Vamos a descartar, de momento, las filas cuarta en adelante

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 3 & 11 & 26 & 4 & 6 \\ 0 & 19 & 21 & 18 & 13 & 8 \end{array}\right) \text{ mod } 27$$

Movemos la tercera fila a segunda fila

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ \mathbf{0} & \mathbf{19} & \mathbf{21} & \mathbf{18} & \mathbf{13} & \mathbf{8} \\ 0 & 3 & 11 & 26 & 4 & 6 \end{array}\right) \text{ mod } 27$$

- Multiplicamos la fila 2 por el inverso de 19 en módulo 27, que es 10
- 0*10 mod 27 = 0 19*10 mod 27 = 1 21*10 mod 27 = 21
- 18*10 mod 27 = 18 13*10 mod 27 = 22 8*10 mod 27 = 26

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ \mathbf{0} & \mathbf{1} & \mathbf{21} & \mathbf{18} & \mathbf{22} & \mathbf{26} \\ 0 & 3 & 11 & 26 & 4 & 6 \end{array}\right) \text{ mod } 27$$

Poniendo 0 en la 3ª fila de la 2ª columna

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ \mathbf{0} & \mathbf{1} & \mathbf{21} & \mathbf{18} & \mathbf{22} & \mathbf{26} \\ \mathbf{0} & \mathbf{3} & \mathbf{11} & \mathbf{26} & \mathbf{4} & \mathbf{6} \end{array}\right) \text{ mod } 27$$

Como ya tenemos un cero en la primera fila de la segunda columna, hacemos:
3ª fila =
3ª fila – 3 * 2ª fila

- 0 – 3*0 mod 27 = 0 3 – 3*1 mod 27 = 0 11 – 3*21 mod 27 = 2
- 26 – 3*18 mod 27 = 26 4 – 3*22 mod 27 = 19 6 – 3*26 mod 27 = 9

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 1 & 21 & 18 & 22 & 26 \\ 0 & 0 & 2 & 26 & 19 & 9 \end{array}\right) \text{ mod } 27$$

- Afortunadamente el 2 tiene inverso en módulo 27, por lo que no hay que trabajar con más filas
- Si ahora multiplicamos la fila 3 por el inverso de 2 en 27, que es el número 14, pondremos el 1 que nos falta para el tercer vector unitario (0 0 1)

Forzando el 1 del vector unitario (0 0 1)

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 1 & 21 & 18 & 22 & 26 \\ 0 & 0 & 2 & 26 & 19 & 9 \end{array}\right) \text{ mod } 27$$

- Multiplicamos la fila 3 por inverso de 2 en 27= 14
- 0*14 mod 27 = 0 0*14 mod 27 = 0
- 2*14 mod 27 = 1 26*14 mod 27 = 13
- 19*14 mod 27 = 23 9*14 mod 27 = 18

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 2 & 11 & 1 & 19 \\ 0 & 1 & 21 & 18 & 22 & 26 \\ 0 & 0 & 1 & 13 & 23 & 18 \end{array}\right) \text{ mod } 27$$

- Si en vez de ese número 2 hubiese aparecido un número con factores en común con el módulo 27, habría que usar alguna fila más de las que antes habíamos descartamos

- Sólo nos falta poner los ceros de la tercera columna (valores 2 y 21) para tener la matriz identidad a la izquierda y terminar así el ataque

Poniendo 0 en la 1ª y 2ª fila de 3ª columna

$$\left(\begin{array}{ccc|ccc} \mathbf{1} & \mathbf{0} & \mathbf{2} & \mathbf{11} & \mathbf{1} & \mathbf{19} \\ 0 & 1 & 21 & 18 & 22 & 26 \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{13} & \mathbf{23} & \mathbf{18} \end{array}\right) \text{ mod } 27$$

- Hacemos
- 1ª fila = 1ª fila – 2 * 3ª fila

- 1 – 2*0 mod 27 = 1 0 – 2*0 mod 27 = 0 2 – 2*1 mod 27 = 0
- 11 – 2*13 mod 27 = 12 1 – 2*23 mod 27 = 9 19 – 2*18 mod 27 = 10

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 12 & 9 & 10 \\ \mathbf{0} & \mathbf{1} & \mathbf{21} & \mathbf{18} & \mathbf{22} & \mathbf{26} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{13} & \mathbf{23} & \mathbf{18} \end{array}\right) \text{ mod } 27$$

- Hacemos
- 2ª fila = 2ª fila – 21 * 3ª fila

- 0 – 21*0 mod 27 = 0 1 – 21*0 mod 27 = 1 21 – 21*1 mod 27 = 0
- 18 – 21*13 mod 27 = 15 22 – 21*23 mod 27 = 25 26 – 21*18 mod 27 = 26

Resultado final con la primera matriz como identidad

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 12 & 9 & 10 \\ 0 & 1 & 0 & 15 & 25 & 26 \\ 0 & 0 & 1 & 13 & 23 & 18 \end{array}\right)$$

Fin del ataque de Gauss-Jordan

- Como la matriz de la izquierda es la matriz de identidad, que cuenta con los tres vectores unitarios 3x3, tomamos su traspuesta cambiando filas por columnas y obtenemos la matriz clave (K) usada en la cifra

$$\begin{pmatrix} 12 & 9 & 10 \\ 15 & 25 & 26 \\ 13 & 23 & 18 \end{pmatrix} \quad \text{Matriz traspuesta} \Rightarrow \quad \text{Clave (K)} \quad \begin{pmatrix} 12 & 15 & 13 \\ 9 & 25 & 23 \\ 10 & 26 & 18 \end{pmatrix}$$

- El criptograma se descifrará utilizando la clave inversa (K^{-1})

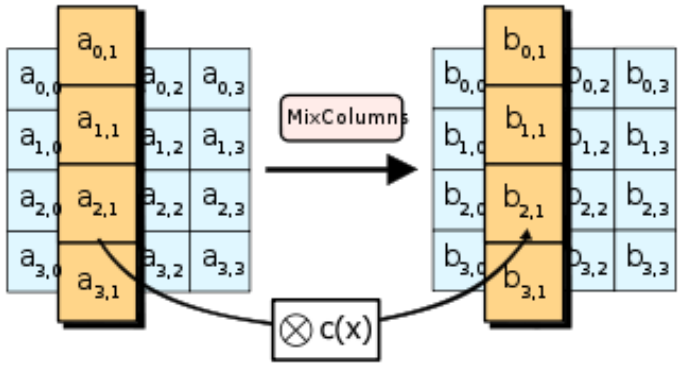
$$\text{Clave } (K^{-1}) \quad \begin{pmatrix} 22 & 22 & 16 \\ 22 & 4 & 24 \\ 25 & 0 & 24 \end{pmatrix}$$

Gauss-Jordan con Criptoclásicos v2.1

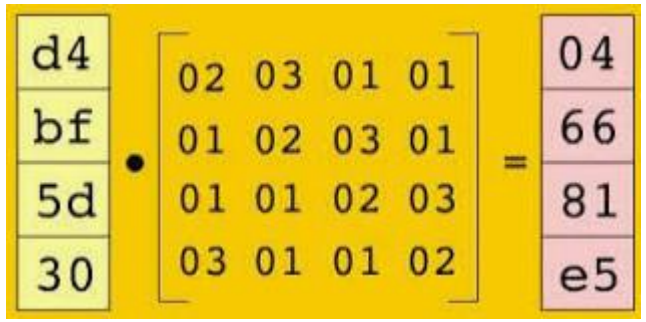
- Criptoanalizar el criptograma C de la cifra de Hill 3x3 en mod 27
- C = NLT GMW IQD OXD OBI GJT PFO OMV BIN PAM ÑTQ ZGK
  1. Conociendo que el texto en claro comienza por LA VIDA ES
  2. Conociendo que el texto en claro comienza por LA VIDA ES COMO
- Criptoanalizar el criptograma C de la cifra de Hill 4x4 en mod 191, sabiendo que el texto en claro comienza por Yo he visto cosas que
- C = ò:¼ä: 4ø½äm a!½´+ +|t{? ís¼^; NÃw£× E®üJ¹ îÅ$Ð$ ÀÕôï² qAh‘5 K"ñC7 5ñ!ÊÅ Ñê7&# ø'Ç-Ï ·ù¿N¸ $ë¿æ» p’R‘ ëß¬öb !÷s'Ø !¯GE= »Á7q^ `ZÕ@ö é'cBf ÅOq{ó î¸Eè¹ Ó²utl Öa<@s ¶j«8Î ¨Íª$ª UGÚ&* ¼éÒN9 pQ2;È ¨¨Mbà ûÛw[' sì¦Ê} ÓkÒÚÛ £ÞLÒÜ êlfa7 i®M‘Ð N^µFþ ®rC>@ E½º

¿No interesan las matrices en la cifra?

- En absoluto, todo lo contrario. Son muy utilizadas para lograr una amplia difusión, pero no de forma única y simple como en Hill sino como parte de otras operaciones no lineales
- Ejemplos de uso:
  - La función MixColumns del algoritmo AES
  - La función *g* en Twofish
- 1999: Algoritmo Cayley-Purser de Sarah Flannery, supuesta competencia a RSA usando matrices, pero sin recorrido



02	03	01	01
01	02	03	01
01	01	02	03
03	01	01	02



Más información en píldoras Thoth

Píldora 22: ¿Cómo se ataca por Gauss-Jordan la cifra de Hill?

https://www.youtube.com/watch?v=7dWtImwcInQ

Conclusiones de la Lección 6.9

- La cifra con matrices de Hill posee un espacio de claves inmenso, pero es muy vulnerable a un ataque con texto en claro conocido
- Buscando (o forzando) los vectores unitarios en el texto en claro o en el criptograma, es posible romper la clave y descifrar el criptograma
- La técnica conocida como Eliminación de Gauss-Jordan para encontrar matrices inversas, nombre en honor a Johann Karl Friedrich Gauss y Wilhelm Jordan, fuerza la aparición de vectores unitarios que permiten romper Hill conociendo tan solo una pequeña parte del texto en claro
- Para romper la cifra de Hill, en muchos casos, sólo nos hará falta tener tantos caracteres del texto en claro y sus correspondientes caracteres del criptograma, como número de elementos tenga la matriz clave

Lectura recomendada

- Guion píldora formativa Thoth nº 22 ¿Cómo se ataca por Gauss-Jordan la cifra de Hill?, Jorge Ramió, 2015
  - https://www.criptored.es/thoth/material/texto/pildora022.pdf
- Libro electrónico de seguridad informática y criptografía, versión 4.1, Jorge Ramió, 2006, archivo Criptografía clásica (Word y PDF), 105 páginas
  - https://www.criptored.es/guiateoria/gt_m001a.htm
- Eliminación de Gauss-Jordan; Wikipedia
  - https://es.wikipedia.org/wiki/Eliminaci%C3%B3n_de_Gauss-Jordan
- Sarah Flannery y su criptosistema de clave pública, Hispasec, 1999
  - https://unaaldia.hispasec.com/1999/12/sarah-flannery-y-su-criptosistema-de-clave-publica.html