Esquemas de cifrado monoalfabéticos

José Galaviz

Cifrado de César

- Se pone el alfabeto en correspondencia con él mismo pero desplazado un cierto número de caracteres.
- Si el alfabeto tiene *n* símbolos, esto define *n* diferentes mapeos (el tamaño del espacio de búsqueda).
- Formalmente hablando el caracter en la posición i del alfabeto se reemplaza, en el criptograma, por el caracter en la posición: $f(i) = i + k \ (mod\ n)$; donde k es el tamaño del desplazamiento.
- El tamaño del desplazamiento es la clave.
- Para n=26 se tienen 25 opciones más la identidad.

a	b	c	d	e	f	g	h	i	j	k	l	m
D	E	F	G	H	I	J	K	L	M	N	O	P
n	o	p	q	r	s	t	u	v	w	x	y	z
Q	R	S	T	U	V	W	X	Y	Z	A	B	C

Alfabeto decimado (diezmado)

- Se pone el alfabeto en correspondencia con él mismo, pero multiplicando en vez de sumando.
- Formalmente hablando el caracter en la posición i del alfabeto se reemplaza, en el criptograma, por el caracter en la posición: $f(i) = i * k \ (mod\ n)$; donde k es el factor de desplazamiento.
- El factor es la clave.
- No cualquier factor es útil: k = 2 usa sólo las posiciones pares y por tanto termina repitiendo letras: f(a) = f(b) con a distinto de b, por lo que e inútil.
- En general k debe ser primo relativo con n.
- Hay menos opciones que en el cifrado de César: para n = 26 se tienen, por ejemplo 1, 3, 5, 7, 9, 11, 15, 17, 19, 21. Doce opciones en vez de las 26 (25 de facto, de César).

Factor 2

0	1	2	3	4	5	6	7	8	9	10	11	12
a	b	c	d	e	f	g	h	i	j	k	l	m
A	C	E	G	I	K	M	O	Q	S	U	W	Y
13	14	15	16	17	18	19	20	21	22	23	24	25
n	o	p	q	r	s	t	u	v	w	x	y	z
A	C	E	G	I	K	M	O	Q	S	U	W	Y

Factor 5

0	1	2	3	4	5	6	7	8	9	10	11	12
a	b	c	d	e	f	g	h	i	j	k	l	m
A	F	K	P	U	Z	E	J	O	T	Y	D	I
13	14	15	16	17	18	19	20	21	22	23	24	25
n	o	p	q	r	s	t	u	v	w	x	y	z
N	S	X	C	H	M	R	W	B	G	L	Q	V

Cifrado afín

- Combinación de los previos: $f(i) = k_1 * i + k_2 \ (mod\ n)$.
- El factor debe cumplir las restricciones del decimado.
- Con n=26, hay 12*26=312 posibilidades, de las cuales una es la identidad.
- Ejemplo:

  $f(i) = 5*i + 2 \ (mod\ 26)$

i	f(i)	i	f(i)
0	2	**13**	15
1	7	**14**	20
2	12	**15**	25
3	17	**16**	4
4	22	**17**	9
5	1	**18**	14
6	6	**19**	19
7	11	**20**	24
2	16	**21**	3
9	21	**22**	8
10	0	**23**	13
11	5	**24**	18
12	10	**25**	23

Afín f(i) = 5 * i + 2 (mod 26)

0	1	2	3	4	5	6	7	8	9	10	11	12
a	b	c	d	e	f	g	h	i	j	k	l	m
C	H	M	R	W	B	G	L	Q	V	A	F	K
13	14	15	16	17	18	19	20	21	22	23	24	25
n	o	p	q	r	s	t	u	v	w	x	y	z
P	U	Z	E	J	O	T	Y	D	I	N	S	X

Alfabeto mezclado

- El caso más general es cuando se permuta arbitrariamente el alfabeto.
- Con n=26 se tiene un espacio de búsqueda de 26! posibilidades: del orden de 10^{26} mapeos.

Sistemas monoalfabéticos

- A los sistemas previos se les denomina monoalfabéticos: el disfraz de cada letra es único y permanente.
- Los cifrados monoalfabéticos preservan la distribución de frecuencias del idioma original.

Criptoanálisis de sistemas monoalfabéticos

1. Sin embargo el criptoanalista resuelve todos los esquemas previos del mismo modo: buscando deducir la correspondencia con base en las estadísticas del idioma original del texto claro. **Análisis de frecuencias**.
2. La tabla de frecuencias, de digramas, de contactos, etc. Son muy útiles.
3. Elementos de un lenguaje (idioma), invariantes bajo cifrados monoalfabéticos.
   a. Distribución de frecuencias.
   b. Palabras frecuentes por tamaño.
   c. Terminaciones, inicios y conectivos frecuentes.
   d. Contactos.
4. Al estilo de *The Gold-Bug* de Edgar Allan Poe (1843).

Inglés

Letra	Frecuencia %	Letra	Frecuencia %
E	12.70	M	2.41
T	9.06	W	2.36
A	8.17	F	2.23
O	7.51	G	2.02
I	6.97	Y	1.97
N	6.75	P	1.93
S	6.33	B	1.49
H	6.09	V	0.98
R	5.99	K	0.77
D	4.25	J	0.15
L	4.03	X	0.15
C	2.78	Q	0.10
U	2.76	Z	0.07

Español (alfabeto de 26 letras, sin ñ no acentos)

Letra	Frecuencia %	Letra	Frecuencia %
E	13.06	P	2.32
A	12.43	B	1.73
O	9.96	H	1.12
S	7.28	Y	1.10
N	7.02	V	1.07
I	6.47	Q	1.06
R	6.46	G	1.06
L	5.92	F	0.58
D	4.78	J	0.54
T	4.30	Z	0.45
U	4.11	X	0.10
C	4.10	K	0.04
M	2.91	W	0.02

Ejemplos

- En español.
  - Si hay palabras de dos letras, una es consonante y otra vocal.
  - Es raro que haya dos consonantes juntas en palabras cortas.
  - Es raro que haya secuencias de consonantes.
  - Las vocales preceden una gran variedad de letras.
  - Las consonantes preceden a unas pocas letras.
  - La Q siempre va seguida de la U y luego de E o I.

Ejemplos

- En español.
  - El 46% de las letras son vocales.
  - Las consonantes N,R y S acumulan el 22% de la frecuencia.
  - 7 letras EAOINRS, acumulan el 65%.
  - PCDESA son frecuentes al inicio de una palabra.
  - EN, ES, ON son frecuentes al derecho y al revés.
  - DE, LA, EL, EN, ES son frecuentes.
  - QUE, LOS, UNA, POR
  - OASEN al final de palabra.

Lo que se preserva

- Luego de cifrar usando un sistema monoalfabético se preserva la distribución de frecuencias de aparición de cada símbolo, salvo el orden.
- Buscar el orden adecuado es más sencillo que explorar todas las permutaciones si la muestra es suficientemente grande (representativa) del idioma en que se escribió el texto claro.
- Cuanto mayor sea la muestra (criptograma), mejor para el criptoanálisis. No hay regla general porque depende de la distribución del idioma, pero suele ser mínimo de 3 a 5 veces el tamaño del alfabeto.

La clave del criptoanálisis

- En el criptograma se preserva algo de la información presente en el texto claro del que proviene.
- En general, uno de los mecanismos del criptoanálisis es encontrar lo que se mantiene invariante luego del cifrado.