

Esquemas de cifrado polialfabéticos

José Galaviz

Romper lo que se preserva

- La clave del criptoanálisis monoalfabético era que se preserva la distribución de frecuencias.
- Porque cada letra tiene un único disfraz.
- Múltiples disfraces, varias opciones por cada letra.
- El número de opciones proporcional a la frecuencia de la letra: cifrado con *homófonos*.
 - Se necesita mayor volumen para el criptoanálisis.
 - Plausible porque no se puede andar cambiando tan rápido.

E.A. Poe otra vez.

- En 1841 E. A. Poe publicó una serie de artículos en Graham's Magazine en los que descifraba criptogramas enviados por los lectores.
- No pudo con tres: dos enviados por W.B. Tyler y uno por G.W. Kulp.
- El primero de Tyler es monoalfabético pero pone al revés cada palabra. Se resolvió en 1991.
- El segundo de Tyler usa varias substituciones posibles por cada letra clara (homófonos). Se resolvió en 2000.
- El de Kulp es cifrado de Vigenere. Estaba publicado con algunos errores. Se descifró en 1975. La palabra clave era *UnitedStates*.

Blaise de Vigenère

- *Traicté des Chiffres et Secretes Manieres d'écrire* (1586).
- Cambia el alfabeto de cifrado con cada letra del mensaje.
- El alfabeto usado para cifrar cada letra lo define una palabra clave.

Cifrado de Vigenère

- Se construye la tabla de cifrado.
- Se empata la palabra clave con el mensaje.
- Se usan las parejas empatadas para determinar, con la tabla, la letra del mensaje cifrado.
- En el esquema original de Vigenere se usaba un alfabeto mezclado que rotaba en cada renglón sucesivo (cifrado de Alberti).

A	C	E	H	J	L	M	N	O	R	S	T
T	A	C	E	H	J	L	M	N	O	R	S
S	T	A	C	E	H	J	L	M	N	O	R
R	S	T	A	C	E	H	J	L	M	N	O
O	R	S	T	A	C	E	H	J	L	M	N
N	O	R	S	T	A	C	E	H	J	L	M
M	N	O	R	S	T	A	C	E	H	J	L
L	M	N	O	R	S	T	A	C	E	H	J
J	L	M	N	O	R	S	T	A	C	E	H
H	J	L	M	N	O	R	S	T	A	C	E
E	H	J	L	M	N	O	R	S	T	A	C
C	E	H	J	L	M	N	O	R	S	T	A

<i>clave</i>	H	O	L	A	H	O	L	A	H	O	L	A	H	O	L	A	<i>renglón</i>
	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	↓	
<i>mensaje original</i>	E	L	M	E	N	S	A	J	E	S	E	C	R	E	T	O	<i>columna</i>
	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	↑	
<i>mensaje cifrado</i>	L	C	T	E	S	M	L	J	L	M	N	C	A	S	J	O	<i>entrada</i>

Análisis de frecuencias, ya no...

- Se tiende a uniformizar la distribución de frecuencias típica del idioma.
- Se pierde el punto de partida del análisis de frecuencias.

Pero...

- La palabra clave es corta, comparada con el texto que se cifra, en general.
- Si no nadie la recordaría.
- Y se repite muchas veces hasta cubrir el texto.
- Debe haber coincidencias.
- Letras que se cifran empatándose con las mismas letras de la clave.

Friederich Wilhelm Kasiski

- Mayor del 33 regimiento de infantería prusiano.
- En 1863 publicó: *Die Geheimschriften und die Dechiffrierkunst* ("La escritura secreta y el arte del descifrado").
- Criptoanálisis de sistemas polialfabéticos.
- Charles Babbage lo descubrió independientemente en 1854, pero sólo quedó registro en su correspondencia personal.

Método de Kasiski

- Buscar coincidencias, cuanto más largas mejor, es menos probable que sean accidentales.
- El texto tiene una temática, así que debe haber palabras que se repiten.
- Observar las distancias entre repeticiones.
- Las distancias entre repeticiones deben ser múltiplos del tamaño de la palabra clave.
- Encontrar el mcd.
- La longitud de la clave debe ser ese o un múltiplo primo de él, la distancia más frecuente, probablemente.
- Dividir en bloques el criptograma, el tamaño del bloque es el tamaño estimado de la clave.
- La primera letra de cada bloque es un monoalfabético de César, la segunda otro y así sucesivamente.

Ejemplo

Repeticiones, distancias y divisores de las distancias			
Secuencia	Frec.	Distancias	Factores primos
WW	4	30, 78, 148	2, 3, 5 / 2, 3, 13 / 2, 2, 37
BWUFEMAK	2	48	2, 2, 2, 2, 3
NH	3	148	2, 2, 37
LIDUWIUS	2	42	2, 3, 7
IAFRXAQEK	2	36	2, 2, 3, 3
YNGO	3	30	2, 3, 5
JEASSVCEATG	2	18	2, 3, 3
NBEKCCM	2	114	2, 3, 19
AMU	2	54	2, 3, 3, 3
MVEFCYNQIERYN	2	18	2, 3, 3
GO	2	60	2, 2, 3, 5

Ejemplo

- En el ejemplo, los factores más frecuentes son 2 y 3, ambos muy pequeños para ser la longitud de la clave, pero 6 podría ser buena idea.
- Luego hay que hacer criptoanálisis de frecuencias en cada columna de los bloques.
- A veces las frecuencias no empatan perfectamente (la segunda o tercera más frecuente es la “e”, p. ej).

William Friedman

- Llegó de un año de edad a EEUU proveniente de Kishniev (Rusia) con su familia en 1892.
- Su nombre original era Wolfe, que cambio por William Frederick.
- Estudió genética en Cornell.
- Comenzó a trabajar en la división de genética de los laboratorios Riverbank de George Fabyan en 1914.
- En la división de criptoanálisis trabajaba Elizebeth Smith analizando la obra de Shakespeare.
- Se casaron en 1921 y fueron a trabajar al Depto. de Guerra.

Elementos

- Aplanar la distribución: si el alfabeto es de 26 letras, que la probabilidad de cada letra sea $1/26$.

$$\sum_{i=1}^{26} \left(p_i - \frac{1}{26} \right)^2 .$$

$$\begin{aligned}
\sum_{i=1}^{26} \left(p_i - \frac{1}{26} \right)^2 &= \sum_{i=1}^{26} \left[p_i^2 - \frac{2}{26} p_i + \left(\frac{1}{26} \right)^2 \right] \\
&= \sum_{i=1}^{26} p_i^2 - \frac{2}{26} \sum_{i=1}^{26} p_i + \sum_{i=1}^{26} \left(\frac{1}{26} \right)^2 \\
&= \sum_{i=1}^{26} p_i^2 - \frac{2}{26} (1) + 26 \left(\frac{1}{26} \right)^2 \\
&= \sum_{i=1}^{26} p_i^2 - \frac{1}{26}
\end{aligned}$$

la suma es la probabilidad de que al elegir con reemplazo dos letras del texto, sean la misma

En la práctica

- Sólo tenemos una muestra finita de texto de tamaño N , donde hay n_i representantes de la i -ésima letra del alfabeto (de 26 letras).
- Esto define el *índice de coincidencias* sobre una muestra.

$$\text{IC}(T) = \sum_{i=1}^{26} \frac{n_i(n_i - 1)}{N(N - 1)}$$

Característico de un idioma

Aproximaciones (usando alfabeto de 26 letras)

- Texto aleatorio: 0.038
- En español: 0.0744
- En inglés: 0.065

Co esto...

- Se puede calcular el número estimado de alfabetos (mapeos) diferentes usados para cifrar un criptograma. Suponiendo texto (T) de tamaño N, en español, representativo, alfabeto de 26 letras.

$$\frac{(0.036)N}{\text{IC}(T)(N - 1) - (0.038)N + 0.0744}.$$

Auxiliar

- Ahora se puede hacer una hipótesis mejor informada acerca del número de mapeos usados en un cifrado polialfabético.
- Es un indicio más que puede usarse junto con la prueba de Kasiski.

Dos veces...

- Se usa al principio para determinar el mejor número de columnas en las que dividir el criptograma.
- Una vez dividido en bloques y columnas el criptograma, el cálculo del índice de coincidencias en cada columna debe ser cercano al del idioma original, lo que confirmaría o refutaría la hipótesis del tamaño de clave.