Sistemas Monoalfabéticos

Este documento no pretende ser un curso exhaustivo, en el mejor de los casos, únicamente puede considerarse como un conjunto de notas complementarias en alguna asignatura. Por supuesto, todo documento es susceptible de mejora, cualquier sugerencia o comunicación de error u omisión será bienvenida.

Cifrado monoalfabético

Las primeras aproximaciones al cifrado de mensajes consideraban una transformación fija para todos los elementos del alfabeto considerado para elaborar el mensaje a cifrar. Todos estos sistemas se conocen como sistemas *monoalfabéticos* y, pese a las diferencias que puedan establecerse respecto el tamaño del espacio de claves o la función de transformación, todos estos métodos pueden atacarse considerando exclusivamente criptograma.

Cifrado por desplazamiento

Una de las primeras aproximaciones documentadas es el *cifrado Caesar* que transformaba cada letra del mensaje sustituyendola de forma cíclica por la letra que ocupaba tres posiciones a su derecha en el el alfabeto latino. Este método puede extenderse para que permita el uso de una clave k, de modo que, si consideramos un alfabeto de n símbolos (habitualmente puede reducirse al conjunto de 27 letras en el alfabeto español en mayúsculas), los ordenamos y codificamos de acuerdo con los valores de $\mathbb{Z}_n$, podemos describir el *cifrado por desplazamiento* como:

$$e_k(x) = x + k \text{ mód } n$$

donde el descifrado puede obtenerse deshaciendo el desplazamiento utilizado.

$$d_k(y) = y - k \text{ mód } n$$

Ejemplo 1. *Dado el mensaje:*

`EJEMPLODEMENSAJEACIFRAR`

el cifrado con la clave $k = 1$ da como resultado:

`FKFNQMPEFNFÑTBKFBDJGSBS`

el cifrado con la clave $k = 6$ da como resultado:

KOKRVQUJKRKSYGOKGIÑLXGX

mientras que si se considera la clave $k = 15$ se obtiene:

SXSAEZDRSASBHOXSOQWTGOG

El *espacio de claves* o número de claves posibles es por lo tanto el tamaño del alfabeto. Obviamente, si se conoce el orden que ocupan los símbolos en el alfabeto, dado que el número de claves habitualmente reducido, el método es sensible a ataques por fuerza bruta.

Cifrado afín

Una forma de, en principio, mejorar la seguridad de un cifrado por desplazamiento es considerar no únicamente una componente de desplazamiento, añadiendo un factor multiplicativo. De este modo, la clave consiste en un par $k = \langle a, b \rangle$, y el cifrado se obtiene como:

$$e_{a,b}(x) = ax + b \text{ mód } n$$

donde n es el tamaño del alfabeto, mientras que el descifrado implica deshacer ambas transformaciones:

$$d_{a,b}(y) = a^{-1}(y - b) \text{ mód } n$$

Ejemplo 2. *Dado el mensaje:*

MENSAJEEJMPLOPARACIFRARCONMETODOAFIN

el cifrado con la clave $\langle a = 7, b = 13 \rangle$ y el alfabeto español en mayúsculas ($n = 27$), el cifrado afín da como resultado:

PÑWLNVÑÑVPQJKQNENAOUENEAKWPÑRKHKNUOW

el cifrado con la clave $\langle a = 11, b = 2 \rangle$ da como resultado:

ZSKVCTSSTZPOFPCLCXJDLCLXFKZSGFIFCDJK

Obviamente, si se considera $a = 1$, el cifrado afín es equivalente a un cifrado por desplazamiento.

Al necesitar el inverso de a para el producto módulo n, únicamente podrá utilizarse factores multiplicativos para los que exista este inverso, por lo que es necesario que $mcd(a, n) = 1$ (es necesario que a sea *relativamente primo* respecto n) y el espacio de claves es igual a $\phi(n) \cdot n$, donde, considerando la descomposición de n en k factores primos:

$$n = \prod_{i=1}^{k} p_i^{e_i}$$

la *phi* de Euler calcula el número de valores de $\mathbb{Z}_n$ relativamente primos respecto n:

$$\phi(m) = \prod_{i=1}^{n} (p_i^{e_i} - p_i^{e_i - 1})$$

Cifrado por sustitución simple

El método de cifrado por sustitución simple no se considera una función concreta que devuelve, dado un elemento del mensaje, el elemento a incluir en el criptograma. En este caso la función viene dada por una de las posibles permutaciones sobre el alfabeto de elementos posibles. De este modo, dada una permutación π, el cifrado considera π como clave, tanto para el cifrado:

$$e_k(x) = \pi(x)$$

como para el descifrado:

$$d_k(y) = \pi^{-1}(y)$$

Ejemplo 3. *Dado el mensaje:*

`VAMOSAENVIARESTEMENSAJECIFRADOPORSUSTITUCIONSIMPLE`

el cifrado con la clave:

```
ABCDEFGHIJKLMNÑOPQRSTUVWXYZ
ÑWALBNGJPUMXDKZCVOYHTISQERF
```

da como resultado:

`SÑDCHÑBKSPÑYBHTBDBKHÑUBAPNYÑLCVCYHIHTPTIAPCKHPDVXB`

El descifrado del mensaje debe considerar de nuevo la permutación pero esta vez en sentido inverso. Como ejemplo, dado el criptograma:

`NEGANENJÑNEXJGVFAFNFNEICYÑVÑXTIÑCSGAMÑVLVAUGNTCFALNFCVTGNEXEGCGXICATECLSJN`

si se sabe que se ha utilizado la clave:

```
ABCDEFGHIJKLMNÑOPQRSTUVWXYZ
VUIFNYMKCBZJLTQASDÑEGXRHWPO
```

el descifrado del criptograma da como resultado el mensaje:

`ESTOESELRESULTADODEDESCIFRARUNCRIPTOGRAMAOBTENIDOMEDIANTESUSTITUCIONSIMPLE`

Es interesante notar que en este caso el espacio de claves aumenta drásticamente, siendo igual al número de permutaciones posibles sobre el alfabeto, igual a 27! si nos limitamos al alfabeto de mayúsculas en español. Pese a que esto impide ataques por fuerza bruta, ya que:

$$27! \approx 2^{94} \approx 10^{29}$$

y el número de segundos desde (lo que se estima) el inicio del Universo es un número de 64 bits, este sistema de cifrado puede ser atacado eficientemente de la misma forma que se abordan ataques al resto de sistemas monoalfabéticos

Criptoanálisis

El criptoanálisis de todos los sistemas monoalfabéticos considera que, en cada idioma, la distribución de los caracteres en un mensaje inteligible no es uniforme. Si consideramos exclusivamente los caracteres del alfabeto sin signos de puntuación ni números y mensajes en español, el elemento más frecuente es el caracter 'E' seguido del caracter 'A', sumando entre ambos más del 25 % de los símbolos de un mensaje. La frecuencia de aparición se resume en la siguiente gráfica de la Figura 1.

Teniendo en cuenta esta distribución no uniforme y que un cifrado monoalfabético transforma siempre los símbolos del mismo modo, podemos tener en cuenta que la misma distribución de probabilidad aparecerá en el criptograma aunque no con la misma asociación presente en los mensajes en claro.

Disponiendo de un análisis de frecuencias de los caracteres del criptograma, el criptoanálisis de un cifrado por desplazamiento se reduce a calcular el desplazamiento entre el caracter más frecuente en un mensaje en el idioma utilizado y el del caracter más frecuente en el criptograma.

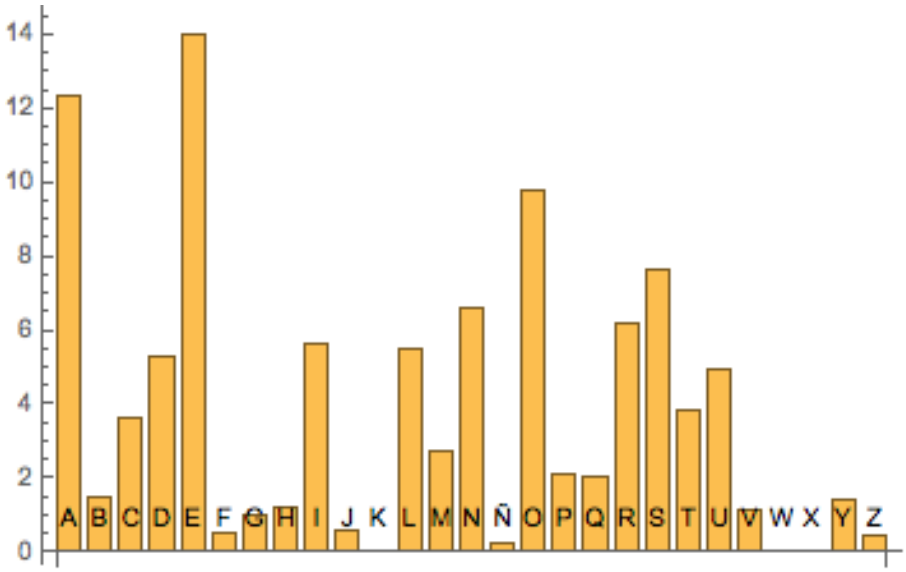

Figura 1: Frecuencia de los símbolos en un mensaje en español.

Ejemplo 4. *Dado el siguiente criptograma obtenido mediante cifrado por desplazamiento:*

`ASBHOXSFJSDQJZIOWBTDGAOQWDBWAEDGIOBISOQSGQORSZDHSXSAEZDH`

un análisis de frecuencias obtiene que el caracter más frecuente es 'S'. Teniendo en cuenta que el caracter más frecuente en español tiene asignado el valor 4 *en* $\mathbb{Z}_{27}$ *(es el quinto del alfabeto) y que 'S' ocupa la posición* 19*, pese a la brevedad del mensaje, hay indicios de que la clave de cifrado pueda ser* 15*. En efecto, considerando esta clave se descifra el mensaje, obteniendo como resultado el mensaje:*

`MENSAJEQUEOCULTAINFORMACIONIMPORTANTEACERCADELOSEJEMPLOS`

En el caso del cifrado afín, el mismo análisis de frecuencias permite hacer una asunción inicial, asociando los caracteres más frecuentes del criptograma a los más frecuentes de un mensaje cualquiera en claro. Esta asunción inicial de pares $\langle caracter\_plano, caracter\_cifrado \rangle$ permite obtener un sistema de dos ecuaciones módulo n cuya resolución proporciona la clave.

Ejemplo 5. *Dado el siguiente criptograma obtenido mediante cifrado afín:*

`VSZVXVESUDVIWCMUORIREVQSWSZVXUUCWEERQROVSIWCMUEUOWV`

Un análisis de frecuencias obtiene que el caracter más frecuente es 'V' con 8 ocurrencias y el segundo más frecuente es 'U' con 6 ocurrencias. Teniendo en cuenta que los símbolos más frecuentes en español son 'E' y 'A', es posible asumir:

$$\begin{cases} 'V' = a'E' + b \text{ mód } 27 \\ 'U' = a'A' + b \text{ mód } 27 \end{cases}$$

Teniendo en cuenta los valores de 'V', 'U', 'E' y 'A' en $\mathbb{Z}_{27}$ (22,21,4,y 0 respectivamente), puede construirse el siguiente sistema:

$$\begin{cases} 22 = 4a + b \text{ mód } 27 \\ 21 = b \text{ mód } 27 \end{cases}$$

Sustituyendo el valor de b en la primera ecuación y resolviendo tenemos:

$$\begin{gathered} 22 = 4a + 21 \text{ mód } 27 \\ 1 = 4a \text{ mód } 27 \\ a = 4^{-1}\, 1 \text{ mód } 27 \\ a = 7 \end{gathered}$$

Una vez obtenida la clave $\langle a = 7, b = 21 \rangle$, la obtención del mensaje oculto se propone como ejercicio.

Por último, el criptoanálisis de un cifrado por sustitución simple supone recuperar la permutación utilizada en el cifrado. De nuevo, un análisis de frecuencias permite hacer una identificación inicial de los caracteres más frecuentes, pero, en cualquier caso, esta asociación permite asociar un máximo de 4 o 5 símbolos, ya que a partir de ese punto, la frecuencia de aparición de los símbolos es muy similar.

A partir de esta asociación inicial es necesario recurrir a los bigramas y/o trigramas más frecuentes, por orden:

es, *ue*, *en*, *de*, *qu*, *os*, *er*, *el*, *as*, *ra*
que, *est*, *ent*, *oqu*, *del*, *con*, *ien*, *ues*, *ade*, *aqu*

Una combinación de estos recursos permite obtener nuevas asociaciones y eventualmente la permutación completa. La casuística es grande, más aún cuando los criptogramas no son suficientemente extensos, por lo que cualquier asociación que se haga debería

corroborarse buscando inconsistencias con el idioma.

Ejemplo 6. *Dado el siguiente criptograma obtenido mediante cifrado por sustitución simple:*

```
JGYLJGNOYJKYLBNJARJYJOSJGJRGNMXTXJOYJPJOYJJKYJOGLAIRIBNJGXRWITLPLJZJPA
HLJONOTXMRISLALRGNGYXYNTXLOGXPAHJAIRIBNJJOJMJTYLGXRWITLPLYIHSJEJRXIRJM
HJZIRBNJTISIGXPELHLIAIRJTJNOONPJRLSJWJTJGGNMNTXJOYJAIRIBNJGJLEGJRWJBNJ
HIMRJTNJOTXISJJGYLGGJIZNGYIIHISJHXSXLPIJGAIDLHNYXHXFISLJOJGYJJZJPAHL
```

Un análisis de frecuencias obtiene el siguiente listado:

$\langle$"J", 53$\rangle$	$\langle$"I", 25$\rangle$	$\langle$"G", 21$\rangle$	$\langle$"L", 21$\rangle$	$\langle$"N", 18$\rangle$
$\langle$"Y", 16$\rangle$	$\langle$"X", 16$\rangle$	$\langle$"R", 16$\rangle$	$\langle$"O", 14$\rangle$	$\langle$"T", 12$\rangle$
$\langle$"H", 11$\rangle$	$\langle$"A", 10$\rangle$	$\langle$"S", 9$\rangle$	$\langle$"P", 9$\rangle$	$\langle$"M", 6$\rangle$
$\langle$"B", 6$\rangle$	$\langle$"Z", 4$\rangle$	$\langle$"W", 4$\rangle$	$\langle$"E", 3$\rangle$	$\langle$"K", 2$\rangle$
$\langle$"F", 1$\rangle$	$\langle$"D", 1$\rangle$	$\langle$"V", 0$\rangle$	$\langle$"U", 0$\rangle$	$\langle$"Q", 0$\rangle$
$\langle$"$\tilde{N}$", 0$\rangle$	$\langle$"C", 0$\rangle$			

Teniendo en cuenta la frecuencia de los símbolos en español, podemos considerar que se han considerado las siguientes asociaciones:

$$\langle'e',"J"\rangle \quad \langle'a',"I"\rangle \quad \langle's',"G"\rangle \quad \langle'o',"L"\rangle$$

(obviamente, podríamos considerar otra asociación para los símbolos 's' y 'o', sin embargo tomaremos esta como la correcta). Teniendo esto en cuenta se obtiene:

```
esYoesNOYeKYoBNeAReYeOSeseRsNMXTXeOYePeOYeeKYeOsoAaRaBNesXRWaToPoeZePA
HoeONOTXMRaSoAoRsNsYXYNTXoOsXPAHeAaRaBNeeOeMeTYosXRWaToPoYaHSeEeRXaReM
HeZaRBNeTaSasXPEoHoaAaReTeNOONPeRoSeWeTessNMNTXeOYeAaRaBNeseoEseRWeBNe
HaMReTNeOTXaSeesYosseaZNsYaaHaSeHXSXoPaesAaDoHNYXHXFaSoeOesYeeZePAHo
```

En este texto podemos distinguir símbolos del texto en claro y del criptograma. Analizando el número de bigramas, uno de los más frecuentes es 'eO' con 9 ocurrencias, que corresponde al bigrama 'en' y que permite encontrar la asociación $\langle'e',"O"\rangle$. Aplicando esta estrategia sucesivamente podríamos eventualmente descifrar todo el texto.

Notamos que, en este caso, la falta de cantidad suficiente de criptograma hace que no se disponga de información acerca de 5 símbolos que no aparecen en el criptograma, haciendo innecesario obtener la clave (permutación) completa para el descifrado del mensaje.

Obviamente, distintos factores como la disponibilidad de cantidad suficiente de criptograma, el autor, el orígen geográfico o la época pueden introducir sesgos que pueden

afectar al proceso. Sin embargo, en la mayoría de los casos, las indicaciones enumeradas permiten reducir la casuística haciendo innecesaria la resolución por fuerza bruta.