

Estimados amigos y amigas:

Este capítulo dedicado a las técnicas de la criptografía clásica, corresponde al tercero del libro de título Aplicaciones Criptográficas, en su segunda edición de junio de 1999, ISBN 83-87238-57-2, publicado por el departamento de Publicaciones de la Escuela Universitaria de Informática de la Universidad Politécnica de Madrid, España.

Puesto que estas técnicas y artilugios no dejan de ser un mero *entretenimiento cultural* en cuanto a su importancia en la criptografía actual, en posteriores ediciones electrónicas de este libro se ha incluido sólo una pequeña reseña al respecto, dándole la importancia que se merecen los sistemas actuales de cifra simétrica y asimétrica, autenticación, firma digital, protocolos, etc. No obstante, y dado que muchas personas me lo han solicitado, lo incluyo también en esta versión 4.0 del libro electrónico, de forma que al lector le sea más fácil estudiar y familiarizarse con estas técnicas de criptoanálisis *ancestrales* ;-) ... y que, sin embargo, son muy didácticas y permiten consolidar muchos conceptos.

Las referencias que encontrará en este documento a capítulos anteriores sobre Teoría de la Información y Matemática Discreta, así como tablas, puede verlas en el libro electrónico: http://www.criptored.upm.es/guiateoria/gt_m001a.htm.

MUY IMPORTANTE: este formato **NO** está optimizado para ninguna impresora. Por tanto, antes de imprimirlo ajuste el formato de este documento a su impresora para que las figuras, tablas y matrices se impriman correctamente.

Madrid, 1 de marzo de 2005
El autor

*Lo que conduce y arrastra al mundo
no son las máquinas sino las ideas.*
Victor Hugo (1802 - 1885)



CRIPTOSISTEMAS CLÁSICOS

Antes de comenzar este capítulo, permítame una breve aclaración. Como habrá podido comprobar por el contenido –si es que lo ha leído claro- al apartado de la criptografía clásica le he dedicado un buen número de páginas, casi la cuarta parte del libro. Si esto es mera historia –*cultura diría alguien y con toda razón*- ¿por qué dedicarle tantas páginas en un libro con un perfil universitario y carácter técnico? Buena pregunta, pero tengo para ello una buena respuesta. Los temas de *Teoría de la Información y Matemática Discreta* vistos en un capítulo anterior pueden entenderse mucho mejor aplicando estos conceptos a criptosistemas reales. Por otra parte, los sistemas que hemos denominado clásicos son especialmente sencillos y didácticos, lo que los convierte en idóneos para fortalecer aquellos conceptos. Además, si este libro cae en manos de un lector o lectora que no tenga este perfil universitario, le será de gran ayuda y seguramente pasará un rato divertido. No hay mucha bibliografía sobre estos temas.

Ahora bien, si no desea *inmiscuirse* en la historia de la criptografía, puede obviar su lectura. No obstante, siempre es bueno adquirir una cierta cultura sobre un tema tan apasionante como éste y que, por una u otra razón, puede estar de moda. Así, en cualquier tertulia podrá hacer el comentario oportuno y ocurrente ante sus amigos/as y *quedar muy bien*. Por lo tanto, aunque sea *a vuelo de pájaro*, le recomiendo su lectura. Además del libro electrónico ya comentado, en la asignatura de Seguridad Informática de la EUI-UPM se ha desarrollado diverso software de prácticas con diversos algoritmos, entre ellos éstos clásicos, entre otros, permitiendo al usuario comprobar en su ordenador las técnicas de cifra, descifrado y, lo que es más interesante, el criptoanálisis de muchos de ellos. Puede encontrarlos junto a otros programas en: <http://www.criptored.upm.es/paginas/software.htm#propio>.

1.1. INTRODUCCIÓN

¿Qué entendemos por criptosistemas clásicos? En un capítulo anterior comentábamos que los sistemas de cifra podían clasificarse de varias formas, siendo la más aceptada aquella que toma en cuenta la característica del secreto de la clave, dando lugar a *criptosistemas de clave secreta* y *criptosistemas de clave pública*. Precisamente en ello nos centraremos tanto en éste como en los siguientes capítulos del libro. Ahora bien, la criptología tal y como hoy en día se concibe, una técnica de *enmascaramiento* de la información estrechamente unida al mundo de la informática, las redes de ordenadores y las autopistas de la información, poco tiene que ver con aquella asociada a fascinantes máquinas de cifrar, que adquirieron gran fama tras su uso en la Segunda Guerra Mundial y más aún, remontándonos a siglos pasados, con los métodos, técnicas y artilugios utilizados por emperadores, gobernantes, militares y en general diversas civilizaciones para mantener sus secretos a buen recaudo.

En aquellos tiempos, el mundo de la criptología estaba vinculado directamente con el *poder fáctico*, ligado a secretos de estado, asuntos militares, de espionaje y diplomáticos, en todo caso siempre seguido de una *aureola de misterio* y que incluso salta a la literatura de ficción en el cuento "*El escarabajo de oro*" de *Edgar Allan Poe*, publicado en 1843 en "*Dollar Newspaper*". Se trata de un relato de aventuras cuyo eje principal gira en torno al *criptoanálisis* de un conjunto de caracteres extraños que aparecen en un pergamino cifrado y cuyo texto esconde el lugar exacto donde se encuentra enterrado el valioso tesoro de un pirata de nombre *Kidd*. El sistema de cifra es uno de los más simples, el denominado *monoalfabético por sustitución con alfabeto mixto*, de forma que nuestro protagonista *William Legrand* no tiene más que aplicar las estadísticas del lenguaje, alguna que otra suposición sobre formación de palabras y una pizca de intuición para hacer corresponder los signos del enigmático criptograma con letras del alfabeto y así describir el mencionado pergamino. Le recomiendo su lectura.

A comienzos del siglo XX el uso de la criptografía en las transmisiones de mensajes cobra una importancia inusitada por los tiempos que corrían (Primera y Segunda Guerras Mundiales), originando esto un gran auge tanto de las técnicas como de las máquinas de cifrar. El 17 de enero de 1917 *William Montgomery*, criptoanalista de la sección diplomática de la famosa Habitación 40 del *Almirantazgo de la Marina Británica* en Londres, intercepta un telegrama lleno de códigos que el Ministro de Relaciones Exteriores alemán *Arthur Zimmermann* envía a su embajador en los Estados Unidos. Tras romper los códigos, descubren atónitos que entre otras cosas el mensaje anunciaba la guerra con los Estados Unidos. Con ello los Estados Unidos entran en la confrontación mundial y ayudan a los aliados a ganar la guerra. Según palabras de *David Khan*, autor de la obra más completa sobre historia de la criptografía¹, "*Nunca un único criptoanálisis ha tenido tan enormes consecuencias*". De hecho, el descubrimiento de este secreto cambió el rumbo de la historia. Y no es el único caso.

¹ Khan, David, "*The Codebreakers. The Story of Secret Writing*", Macmillan Publishing Company, New York, 1967, pp. 266 ss. y 282 ss.

Otro ejemplo histórico lo tenemos en plena Segunda Guerra Mundial. El 7 de diciembre de 1941, la radio de la estación naval de Bainbridge Island, cerca de Seattle, intercepta un mensaje de solamente 9 minutos desde Tokyo a la Embajada Japonesa en los Estados Unidos. El radiotelegrama estaba cifrado con una máquina que los norteamericanos llamaron *Purple*, cuyo código fue roto por *William Friedman*, quizás el criptólogo más importante de la historia, y un grupo de criptoanalistas. Si bien es cierto que ello no pudo evitar el ataque de los japoneses a Pearl Harbor, el esfuerzo realizado por todos en la destrucción de tales códigos jugó luego un papel fundamental y marcó la derrota del pueblo nipón así como el fin de la guerra.

En resumen, si se repasa la historia de la primera mitad del siglo XX y en especial todo lo relativo a la información secreta que se transmitía por radio en forma cifrada y que, tras ser interceptada por el enemigo, era criptoanalizada en verdaderas empresas *rompedoras de códigos*, no resulta nada extraño las afirmaciones hechas por políticos de la época en cuanto a que el uso de las técnicas criptográficas *cambió el curso de los acontecimientos*, desequilibrando la balanza hacia un sentido. El lector interesado en este apasionante tema histórico, encontrará en el libro de *David Khan* "*The Codebreakers*" -verdadero *tratado* sobre la historia de la criptología clásica- una lectura amena y llena de anécdotas sobre las aplicaciones de la criptografía desde sus albores hasta la década de los sesenta. A partir de esta época, serán los ordenadores y la informática quienes toman el relevo del protagonismo en los sistemas de cifra.

Decíamos en un capítulo anterior que dos hechos significativos marcan un punto de inflexión en el mundo de la criptografía. El primero de ellos, los estudios que en el año 1948 realiza *Claude Shannon* sobre teoría de la información y criptología: desde ese momento, la criptología deja de ser considerada como un mero arte rodeado de un cierto aire de misterio y en algunos casos excepticismo, para ser tratada como una rama más de las matemáticas. Hoy también tienen un papel fundamental la informática y las ciencias de la ingeniería. El segundo hecho es la publicación en el año 1976 de un artículo por parte de *Whitfield Diffie* y *Martin Hellman* en el que proponen una nueva filosofía de cifra, dando lugar a los criptosistemas de clave pública.

Según lo anterior, podríamos afirmar entonces que la criptografía clásica abarca desde tiempos inmemoriales, como veremos a continuación, hasta los años de la posguerra, es decir, hasta la mitad del siglo XX. El adjetivo de *clásica*, en contraposición al de criptosistemas *modernos*, se debe tanto a las técnicas utilizadas en las primeras, básicamente operaciones de sustitución y transposición de caracteres, con o sin clave pero siempre unido al concepto de clave secreta, como al uso de máquinas dedicadas a la cifra. En el caso de los sistemas modernos, éstos hacen uso, además de lo anterior, de algunas propiedades matemáticas como, por ejemplo, la dificultad del cálculo del logaritmo discreto o el problema de la factorización de grandes números, como vimos en el capítulo anterior, unido esto a la representación binaria de la información. No obstante, muchos sistemas modernos y que en la actualidad se siguen utilizando, como los algoritmos de clave secreta DES e IDEA, se basan en conceptos que podríamos denominar clásicos como son los de transposición y sustitución con una clave privada, si bien en estos sistemas la operación se realiza sobre una cadena de bits y no sobre caracteres.

Muchos de los criptosistemas clásicos, en particular aquellos que transforman el mensaje en claro aplicando técnicas de sustitución y transposición, basan su seguridad principalmente en el secreto de la transformación o algoritmo de cifra. Es ésta también una diferencia fundamental con respecto a los sistemas modernos, en los que el algoritmo se hace público puesto que la fortaleza del sistema reside en la imposibilidad computacional de romper una clave secreta. Observe que el hacer público el algoritmo de cifra permite al criptólogo evaluar la calidad del software desarrollado, en tanto será estudiado por la comunidad científica intentando buscar un defecto, una *puerta falsa*, una rutina innecesaria, una codificación no depurada, etc.

De todos los sistemas clásicos, cuya diversidad es enorme como puede comprobar el lector del libro de Khan, en este capítulo sólo analizaremos algunos; los más conocidos y que, de alguna forma, nos servirán como apoyo para profundizar y aplicar algunos conceptos que sobre criptosistemas, seguridad informática, teoría de la información, de los números y de la complejidad de los algoritmos han sido estudiados en los capítulos anteriores.

1.1.1. Un poco de historia

- **La escítala**

Ya en siglo V antes de J.C. los lacedemonios, un antiguo pueblo griego, usaban el método de la *escítala* para cifrar sus mensajes. El sistema consistía en una cinta que se enrollaba en un bastón y sobre el cual se escribía el mensaje en forma longitudinal como se muestra en la Figura 1.1.

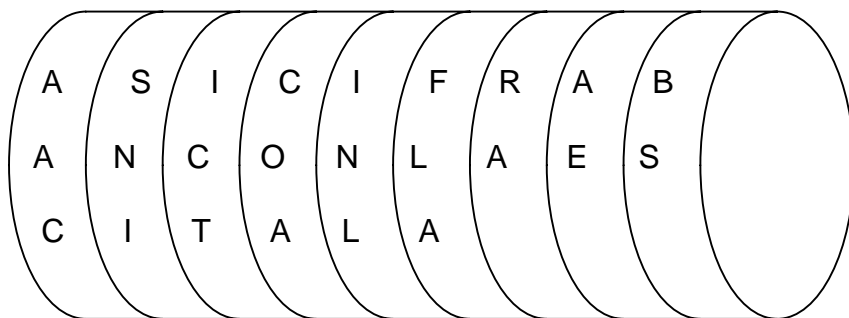


Figura 1.1. Cifrado mediante sistema de escítala.

Una vez escrito el mensaje, la cinta se desenrollaba y era entregada al mensajero; si éste era interceptado por cualquier enemigo, lo único que se conseguía era un conjunto de caracteres o letras distribuidas al parecer de forma aleatoria en dicha cinta. Incluso si el enemigo intentaba enrollar la cinta en un bastón con diámetro diferente, el resultado obtenido era un conjunto de letras escritas una a continuación de otra sin sentido alguno. Por ejemplo, en el caso de la figura 1.1, la cinta llevará el mensaje $M = \text{ASI CIFRABAN CON LA ESCITALA}$ si bien en ella sólo podrá leerse el criptograma $C = \text{AACSNIICTCOAINLFLARAAEBS}$. Para enmascarar completamente la escritura, es obvio que la cinta en cuestión debe tener caracteres en todo su contorno. Como es de esperar, la clave del sistema residía precisamente en el diámetro de aquel

bastón, de forma que solamente el receptor autorizado tenía una copia exacta del mismo bastón en el que enrollaba el mensaje recibido y, por tanto, podía leer el texto en claro. En este sistema no existe modificación alguna del mensaje; es decir, éste va *en claro* desde el transmisor hacia el receptor, por lo que como veremos más adelante se tratará de un cifrador por transposición.

De esta forma se lograba el objetivo de la confidencialidad, en tanto que la integridad estaba en entredicho y dependía de lo *aguerrido* y *fiel* que fuese nuestro mensajero. Si la cinta era robada y se cambiaban los caracteres, podría llegar al receptor un mensaje sin sentido y, lo que es peor, con un duplicado del bastón original podía enviarse un mensaje con sentido completamente distinto al encomendado al mensajero. Haga un viaje mental al pasado e imagínese lo que significaría en aquellos tiempos que el destinatario recibiera el mensaje falso $M_F = \text{RENDICIÓN TOTAL}$ en vez del verdadero mensaje $M_V = \text{ATACAMOS MAÑANA}$, ambos de 14 caracteres. Sin duda a más de alguno este *desliz* le costaría su *preciada cabeza*.

Para terminar, un apunte curioso y de cultura general. De estos tiempos tan remotos se debe la famosa frase de ostentar el "*bastón de mando*" –tan popular entre nuestros queridos políticos y en particular alcaldes– y que, como es de suponer, en aquella época no se soltaba por ningún motivo puesto que en él residía la seguridad del sistema de información y la vida política de este pueblo de la antigua Grecia.

• El cifrador de Polybios

A mediados del siglo II antes de J.C., encontramos el cifrador por sustitución de caracteres más antiguo que se conoce. Atribuido al historiador griego *Polybios*, el sistema de cifra consistía en hacer corresponder a cada letra del alfabeto un par de letras que indicaban la fila y la columna en la cual aquella se encontraba, en un recuadro de $5 \times 5 = 25$ caracteres, transmitiéndose por tanto en este caso el mensaje como un criptograma. En la Figura 1.2 se muestra una tabla de cifrar de Polybios adaptada al inglés, con un alfabeto de cifrado consistente en el conjunto de letras A, B, C, D y E aunque algunos autores representan el alfabeto de cifrado como los números 1, 2, 3, 4 y 5.

| | A | B | C | D | E | | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | A | B | C | D | E | 1 | A | B | C | D | E |
| B | F | G | H | I | K | 2 | F | G | H | I | K |
| C | L | M | N | O | P | 3 | L | M | N | O | P |
| D | Q | R | S | T | U | 4 | Q | R | S | T | U |
| E | V | W | X | Y | Z | 5 | V | W | X | Y | Z |

Figura 1.2. Tablas de cifrar de Polybios.

Acorde con este método, la letra A se cifrará como AA, la H como BC, etc. Esto significa que aplicamos una sustitución al alfabeto $\{A, B, C, \dots, X, Y, Z\}$ de 26 letras convirtiéndolo en un alfabeto de cifrado $\{AA, AB, AC, \dots, EC, ED, EE\}$ de 25 caracteres, si bien sólo existen 5 símbolos diferentes $\{A, B, C, D, E\}$. Este tipo de tabla o matriz de cifrado será muy parecida a la que en el siglo XIX se utilizará en el criptosistema

conocido como cifrador de Playfair y que será tratado más adelante en el apartado de cifradores poligráficos, salvo que en este último la operación de cifra no se realiza por monogramas como en el de Polybios sino por digramas, conjunto de dos caracteres del texto en claro.

Ejemplo 1.1: *Usando la Tabla del cifrador de Polybios, cifre el mensaje:*

M = QUE BUENA IDEA LA DEL GRIEGO.

Solución: *C = DADEAE ABDEAECCAA BDADAEAA CAAA ADAECA
BBDBBDAEBBCD.*

Aunque resulte elemental, se deja como ejercicio para el lector encontrar el criptograma cuando se utiliza la tabla de Polybios con representación numérica. El criptograma que se obtiene con este cifrador tiene una extensión de caracteres igual al doble de la del texto en claro, característica que no puede considerarse precisamente como una *virtud* de este método de cifra. En realidad no fue tan buena la idea.

• El cifrador del César

Unos cincuenta años después del cifrador de Polybios, en el siglo I antes de J.C., aparece un cifrador básico conocido con el nombre genérico de cifrador del César en honor al emperador *Julio César* y en el que ya se aplica una transformación al texto en claro de tipo monoalfabética. Como se verá en un apartado posterior, el cifrador del César aplica un desplazamiento constante de tres caracteres al texto en claro, de forma que el alfabeto de cifrado es el mismo que el alfabeto del texto en claro pero desplazado 3 espacios hacia la derecha módulo n, con n el número de letras del mismo. En la Figura 1.3. se muestra el alfabeto y por tanto la transformación que utiliza este cifrador por sustitución de caracteres para el alfabeto castellano de 27 letras.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M_i | 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_i | 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 |

Figura 1.3. Alfabeto de cifrado del César para lenguaje castellano.

Ejemplo 1.2: *Con el cifrador del César según el alfabeto mostrado en la Figura 1.3, cifre los siguiente mensajes:*

M₁ = VINI, VIDI, VINCI. (Frase célebre de César: Llegué, vi, vencí).

M₂ = AL CÉSAR LO QUE ES DEL CÉSAR.

Solución: *Aplicando a cada carácter M_i su equivalente C_i de la tabla de la Figura 1.3, se obtienen los siguientes criptogramas:*

C₁ = YLPL, YLGL, YLPFL.

C₂ = DÑ FHVDU ÑR TXH HV GHÑ FHVDU.

A partir del ejemplo anterior, es fácil apreciar ciertas *debilidades* en este cifrador como, por ejemplo, la repetición de la cadena de caracteres YL en el criptograma primero y FHVDU en el segundo que entregan demasiadas pistas a un posible criptoanalista. Estos y otros puntos débiles del cifrador del César que por ahora no saltan a la vista serán analizados y comentados más adelante.

• El cifrador de Alberti

En el siglo XVI *Leon Battista Alberti* presenta un manuscrito en el que describe un disco cifrador con el que es posible cifrar textos sin que exista una correspondencia única entre el alfabeto del mensaje y el alfabeto de cifrado como en los casos analizados anteriormente. Con este sistema, cada letra del texto en claro podía ser cifrada con un carácter distinto dependiendo esto de una clave secreta. Se dice entonces que tales cifradores usan más de un alfabeto por lo que se denominan *cifradores polialfabéticos*, a diferencia de los anteriores denominados *monoalfabéticos*.

Como se aprecia en la Figura 1.4, el disco de Alberti presenta en su círculo exterior los 20 caracteres del latín, esto es, los mismos del alfabeto castellano excepto las letras H, J, Ñ, K, U, W e Y, y se incluyen los números 1, 2, 3 y 4 para códigos especiales. Por su parte, en el disco interior aparecen todos los caracteres del latín además del signo & y las letras H, K e Y. Al ser 24 los caracteres representados en cada disco, es posible definir hasta 24 sustituciones diferentes; es decir, dependiendo de la posición del disco interior la cantidad máxima de alfabetos de cifrado es igual a 24. Luego, para cifrar un mensaje, una vez establecida la correspondencia entre caracteres de ambos discos o, lo que es lo mismo, el alfabeto de cifrado, se repasa letra a letra el texto en claro del disco exterior y se sustituye cada una de ellas por la letra correspondiente del disco interior.

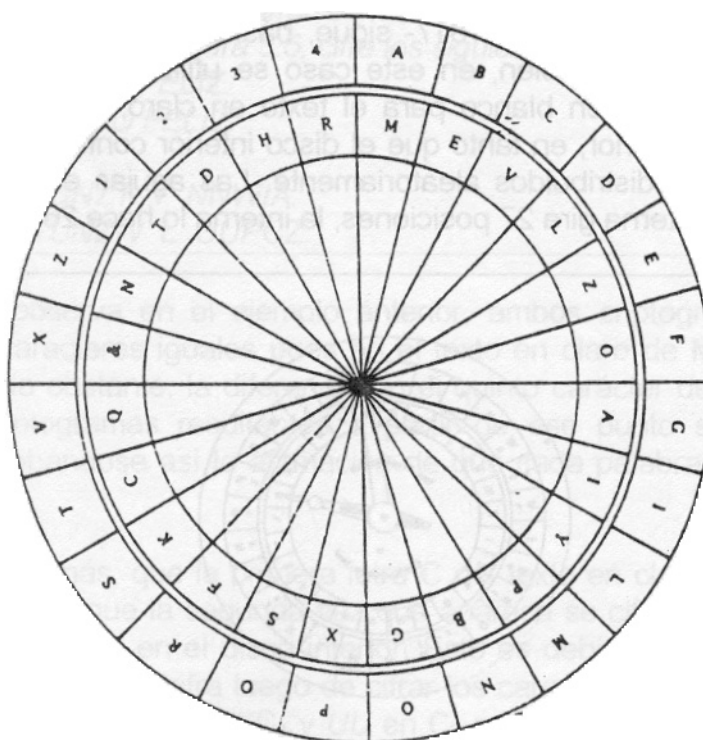


Figura 1.4. Disco cifrador de Alberti.

La innovación que supone este sistema consiste en que el alfabeto de sustitución puede ser cambiado durante el proceso de cifrado, por ejemplo cada k caracteres, simplemente girando el disco interior y por tanto utilizando otro alfabeto de sustitución.

Ejemplo 1.3: *Cifre con el disco de Alberti de la Figura 1.4, siendo su posición inicial la de coincidencia entre el número 1 del disco exterior y el signo & del disco interior, el siguiente mensaje:*

*M = EL DISCO DE ALBERTI ES EL PRIMER CIFRADOR
POLIALFABÉTICO CONOCIDO.*

Solución: *Desplazamos el disco interior dos espacios en el sentido de las agujas del reloj y leemos el carácter cifrado en el disco interior bajo el carácter correspondiente del texto en claro del disco exterior, obteniéndose:*

*C = VA EOSMP EV HARVXFO VS VA BXOIVX MOLXHEPX
BPAOHALHRVFOMP MPYPMOEP.*

1.1.2. CIFRADORES DEL SIGLO XIX

En el siglo XIX comienzan a desarrollarse diversos sistemas de cifra con las características polialfabéticas propuestas por Alberti, entre los que destacan el de discos concéntricos de Wheatstone en 1860 y el de cilindros de Bazeris en 1891.

- **El cifrador de Wheatstone**

El criptógrafo de *Wheatstone* mostrado en la Figura 1.5. -según un invento de *Decius Wadsworth* desarrollado en 1817- sigue, básicamente, el mismo algoritmo de cifra que el de Alberti. Ahora bien, en este caso se utiliza el alfabeto inglés de 26 caracteres más el espacio en blanco para el texto en claro, representado de forma ordenada en el disco exterior, en tanto que el disco interior contiene solamente los 26 caracteres del lenguaje distribuidos aleatoriamente. Las agujas están engranadas de forma que cuando la externa gira 27 posiciones, la interna lo hace 26.

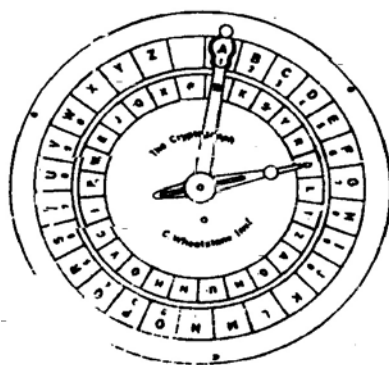


Figura 1.5. Máquina de cifrar de Wheatstone.

El método de cifra consiste en hacer girar la aguja externa en el sentido de las manecillas del reloj hasta hacer coincidir cada letra del texto en claro con la letra del disco externo y apuntar el carácter correspondiente que aparece en el círculo interior, incluso para el espacio en blanco. Observe que por la relación de giro de las agujas, éstas se van separando una posición o letra por cada vuelta, de forma que el alfabeto de cifrado será diferente cuando se cumpla cualquiera de estas tres condiciones:

- a) Que se termine una palabra del texto en claro y por tanto demos un giro completo de la aguja mayor al buscar el espacio en blanco.
- b) Que aparezcan letras repetidas y tengamos que dar toda una vuelta completa al buscar la segunda. No obstante, según los autores, en este caso es posible también omitir cifrar la letra repetida o bien cifrar ambas como una única letra poco usual, por ejemplo la letra Q.
- c) Que las letras de una palabra no vengan en orden alfabético. Es decir, si ciframos la palabra *CELOS* no alcanzamos a dar la vuelta completa al disco exterior, en tanto que la palabra *MUJER* implica dos vueltas y *HOMBRE* significa tres. No trate de encontrar ningún *mensaje subliminal* en estas tres palabras y sus vueltas.

La importancia de este cifrador está en que cada una de las palabras del mensaje influye en la forma en que se cifran las siguientes, una propiedad muy interesante y que precisamente utilizarán los cifradores modernos, sencillamente definiendo el concepto de palabra como bloque de bits para la cifra y aplicando lo que se denomina cifrado con encadenamiento.

Ejemplo 1.4: *Con la máquina de cifrar de Wheatstone y suponiendo la posición inicial indicada en la Figura 1.5, cifre los siguientes mensajes:*

$M_1 = \text{CHICA FELIZ.}$

$M_2 = \text{CHICO FELIZ.}$

Solución: $C_1 = \text{TUNZT T NNWIA.}$
 $C_2 = \text{TUNZW L UUPCZ.}$

Como se observa en el ejemplo anterior, ambos criptogramas presentan los cuatro primeros caracteres iguales pues en el texto en claro de M_1 y M_2 también son iguales (CHIC). No obstante, la diferencia en el quinto carácter de los textos M_1 y M_2 , hace que los criptogramas resultantes a partir de ese punto sean completamente diferentes, comprobándose así la afirmación de que cada palabra influye en el cifrado de la siguiente.

Observe, además, que la primera letra C del texto en claro en ambos casos se cifra como T, en tanto que la segunda vez que aparece se cifra como Z, precisamente un espacio hacia delante en el disco interior. Esto es debido al giro completo que se produce en la operación de cifra luego de cifrar los caracteres C, H e I. Por otra parte, los caracteres repetidos TTNN en C_1 y UU en C_2 se deben a una revolución completa del disco interior producida por dos caracteres contiguos en el texto en claro y que están separados 26 espacios como es el caso de los digramas "A " y "FE". Por último, apréciase que una misma palabra repetida en el texto en claro se cifrará cada vez con un alfabeto distinto por la rotación completa producida por la búsqueda del espacio en blanco.

Por ejemplo el mensaje $M = \text{TORA TORA}$, palabra secreta usada como clave por los japoneses en el ataque a Pearl Harbor y cuyo significado es tigre, se cifrará como $C = \text{XWQT Z KQBG.}$

• El cifrador de Bazeries

El cifrador de *Étienne Bazeries*, criptólogo francés nacido a finales del siglo XIX, está basado en el cifrador de ruedas de Jefferson, inventado unos 100 años antes por *Thomas Jefferson* reconocido como el padre de la criptografía americana. El criptógrafo mostrado en la Figura 1.6 consta de 20 discos, cada uno de ellos con 25 letras en su circunferencia, de forma que la clave se establece sobre la generatriz del cilindro, determinándose 25 alfabetos diferentes. Su funcionamiento es el siguiente: para cifrar el mensaje, primero se divide éste en bloques de 20 letras, procediendo luego a su colocación en forma longitudinal en la línea del visor. El criptograma que se envía puede ser cualquiera de las 25 líneas, también llamadas generatrices del cilindro. Por ejemplo, si se elige la generatriz de distancia +2 en la Figura 1.6, el mensaje $M = \text{JE SUI S INDECHIFFRABLE}$ del visor se cifraría como $C = \text{LOVS PQUU TPUKEJHHCFDA}$.

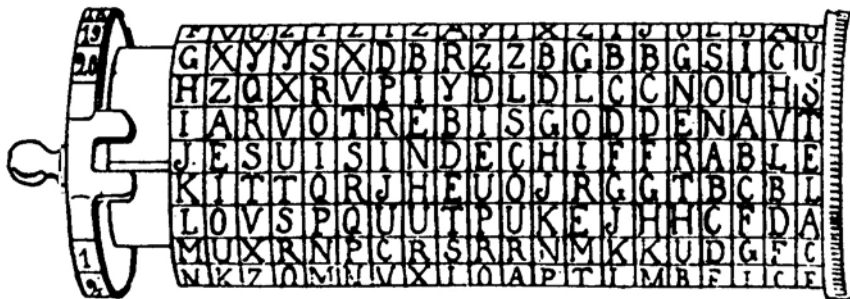


Figura 1.6. Máquina de cifrar de Bazeries.

Se puede elegir la misma distancia a la generatriz en la cual se lee el criptograma para todo el bloque o bien cambiar ésta en cada bloque o elemento del bloque, de forma que el número de combinaciones o alfabetos distintos en vez de ser solamente 25 podría crecer hasta el factorial de 25, un valor verdaderamente alto. Uno de estos posibles alfabetos podría ser elegir una secuencia de distancias, una vez introducido el mensaje en el visor, igual a -1,-2,-2,-1,1,2,2,1,-1,-2,-2,-1,1,2,2,1,-1,-2,-2,-1. Es decir, una vez se tenga el mensaje en claro en el visor, se envía como primer carácter del criptograma el que, en la misma columna, está desplazado una posición hacia atrás en el anillo; como segundo el que está desplazado dos posiciones atrás, el tercero también dos posiciones atrás, el cuarto una posición atrás, el quinto una posición hacia delante, el sexto dos adelante, etc., de manera que el criptograma forma una especie de zig-zag en torno al texto en claro, sin transmitir ningún carácter de éste puesto que la posición 0 no se encuentra en la secuencia indicada. Como es fácil observar, dicha secuencia sería la clave del sistema y, en este caso, su valor máximo sería igual todas las posibles permutaciones es decir $25! = 1,55 \times 10^{25}$, un valor muy grande aunque el sistema de cifra sería engorroso y poco práctico.

La operación de descifrado consiste en poner los caracteres del criptograma en el visor y buscar en alguna de las líneas el mensaje en claro o seguir el proceso inverso al comentado anteriormente. Como los bloques de criptograma tienen longitud de veinte caracteres, es prácticamente imposible que exista más de una solución con sentido.

Ejemplo 1.5: Considerando una representación del cifrador de Bazeries como la que se indica a continuación, cifre el mensaje mostrado en el visor de la generatriz 11 del disco: $M = \text{INTENTA ROMPER LA CIFRA}$.

a) Con una distancia constante de +3 espacios.

b) Con la secuencia S de distancia de cifrado indicada:

$S = 0, 1, 2, 1, 0, -1, -2, -1, 0, 1, 2, 1, 0, -1, -2, -1, 0, 1, 2, 1$.

| Fila | Disco | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |
|------|-------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------|
| | | | | | | | | | | | | | | | | | | | | | |
| 7 | | V | A | M | W | W | U | I | O | P | S | S | A | H | K | L | V | C | D | U | Q |
| 8 | | M | O | J | F | J | K | L | M | A | C | H | Y | E | D | X | Z | G | Q | I | U |
| 9 | | D | B | W | Q | D | S | <u>B</u> | D | Q | T | F | D | S | F | <u>D</u> | X | E | W | Q | G |
| 10 | | A | W | K | Y | H | <u>M</u> | P | <u>E</u> | S | H | U | K | P | <u>U</u> | O | <u>E</u> | S | J | K | D |
| 11 | | <u>I</u> | N | T | <u>E</u> | <u>N</u> | T | A | R | <u>O</u> | M | P | <u>E</u> | <u>R</u> | L | A | C | <u>I</u> | F | R | A |
| 12 | | R | <u>G</u> | I | <u>S</u> | X | F | W | G | B | <u>A</u> | N | <u>L</u> | F | M | J | H | A | <u>A</u> | S | <u>H</u> |
| 13 | | U | I | <u>D</u> | V | C | R | I | I | Z | D | <u>D</u> | C | Z | A | K | I | M | B | <u>L</u> | X |
| 14 | | K | L | B | O | T | Z | H | Y | L | V | C | O | N | D | W | A | L | M | V | Z |
| 15 | | W | H | S | K | L | P | O | U | I | E | E | P | D | N | C | G | Q | E | O | B |

.....

Solución: a) $C = \text{KLBOTZHYLVCONDWALMVZ}$.

b) Según la secuencia indicada, tomamos caracteres consecutivos de las líneas 11, 12, 13, 12, 11, 10, 9, 10, 11, 12, 13, 12, 11, 10, 9, 10, 11, 12, 13, 12 que se encuentran subrayados. $C = \text{IGDSNMBEOADLRUDEIALH}$.

Todos los sistemas comentados y muchísimos otros que se desarrollaron paralelamente en Europa, América y Asia, han sido criptoanalizados incluso sin contar con la ayuda de equipos informáticos. No obstante, la discusión de su criptoanálisis está fuera del objetivo de este libro, por lo que al lector interesado en tales temas históricos se le remite nuevamente al libro de Khan y a las publicaciones que se indican.²

1.1.3. Máquinas de cifrar en el siglo XX

Ya entrado el siglo XX, aproximadamente unos 20 años antes de que estalle la Segunda Guerra Mundial, se desarrollan diversas máquinas de cifrar con rotores o ruedas que permiten un cifrado polialfabético, dando lugar a un importante número de claves secretas que, para aquel entonces, dificultaba *in extremis* el criptoanálisis. Este desarrollo a nivel industrial de los criptosistemas resulta lógico pues en aquellos años previos a dicha confrontación mundial, estaba todavía muy fresco en la memoria de todos, y en especial de gobernantes y militares, los efectos de la Primera Guerra Mundial, por lo que las medidas de seguridad ante el miedo al espionaje adquirían una importancia vital. Recuerde el famoso telegrama de Zimmermann comentado al comienzo del capítulo.

² Deavours, Cipher; Khan, David; Kruh, Louis; Mellen, Greg; Winkel, Brian, "Cryptology: Machines, History & Methods", Artech House, 1989. Bauer, F.L., "Decrypted Secrets. Methods and Maxims of Cryptology", Springer, 1991.

De estas máquinas, cuyo papel principal fue su utilización para enviar mensajes cifrados precisamente en la Segunda Guerra Mundial, destacan tanto por sus características como por el halo de misterio que las rodeaba dos de ellas: la máquina *Enigma* y la de *Hagelin*. Encontrará fotografías y esquemas muy interesantes sobre éstas y otras máquinas de cifrar en el libro de Khan, en la referencia anterior y en el software del Libro Electrónico de Criptografía Clásica en Toolbook que puede descargar gratuitamente desde http://www.criptored.upm.es/software/sw_m001a.htm.

• La máquina Enigma

Inventada por el ingeniero alemán *Arthur Scherbius* en el año 1923, la máquina Enigma consiste en un banco de rotores montados sobre un eje, en cuyos perímetros había 26 contactos eléctricos, uno por cada letra del alfabeto inglés. En realidad el precursor de este tipo de máquinas con rotores fue *Edward Hugh Hebern* que algunos años antes inventa y comercializa los denominados *cifradores de códigos eléctricos*. Esta máquina debe su fama a la amplia utilización durante la Segunda Guerra Mundial, en especial por parte del ejército alemán. El imperio japonés también cifra sus mensajes con una máquina similar denominada *Purple*. Estos códigos, por muy difíciles que puedan parecer, fueron rotos por los criptoanalistas de la época.

Los rotores se desplazan como un odómetro. Es decir, al cifrar un carácter el primer rotor avanza una posición y sólo cuando éste ha realizado una rotación completa, el segundo se desplaza un carácter, y así sucesivamente. Estos volverán a su posición inicial, tras un período igual a n^t . Por ejemplo, en un sistema con 4 rotores, se utilizan de $26^4 = 456.976$ alfabetos. Si aumentamos los rotores a 5, esta cantidad asciende a 11.881.376. La operación de cifra para estas máquinas sigue la siguiente congruencia:

$$E_i(M) = (f_i(M - p_i) \bmod 26 + p_i) \bmod 26 \quad \boxed{1.1}$$

En la ecuación anterior, p_i es la posición en la que se encuentra el rotor i ésimo y f_i la correspondencia de los caracteres de la cara anterior y posterior de este rotor. Por lo tanto, el carácter i ésimo M_i del mensaje $M = m_1m_2m_3\dots$ se cifrará como:

$$E_{ki}(M_i) = F_t * \dots * F_1(M) \quad \boxed{1.2}$$

• La máquina Hagelin

La máquina Hagelin fue inventada por el criptólogo sueco *Boris Hagelin*, quien adquirió en 1927 la fábrica de máquinas de cifrar de *Arvid G. Damm*, otro inventor sueco que no tuvo la suerte de sacar un producto competitivo en el mercado. Entre los años veinte y los treinta, Hagelin diseña diversas máquinas (B-21, B-211, C-36, C-48, etc.) en las que a través de ruedas con piñones realiza una cifra similar a la utilizada por el sistema de Beaufort que veremos más adelante.

La particularidad de estas máquinas que a la postre hizo millonario a Hagelin, probablemente ante la desesperación de Damm, estaba en una periodicidad muy alta puesto que el número de dientes de las diferentes ruedas eran primos entre sí. Para

seis ruedas estos valores eran 26, 25, 23, 21, 19 y 17, de forma que el período era igual a su producto, un valor que supera los 100 millones. La ecuación matemática que representa al cifrado de Hagelin es:

$$E_{k_i}(M_j) = (k_i - M_j) \bmod 26$$

1.3

1.2. ALFABETOS Y CARACTERÍSTICAS DEL LENGUAJE

1.2.1. Alfabetos de cifrado

En la mayoría de los cifradores clásicos se utiliza como alfabeto de cifrado el mismo alfabeto del texto en claro. Para poder aplicar las operaciones de transformación se asocia a cada letra del alfabeto un número de forma que a la letra A le corresponde el 0, a la letra B el 1, etc. De esta manera, si nos centramos en el castellano, podríamos definir en principio cinco tipos de alfabetos:

Alfabeto 1: Letras mayúsculas: aritmética módulo 27.

Alfabeto 2: Letras mayúsculas con números 0-9: aritmética módulo 37.

Alfabeto 3: Letras mayúsculas y minúsculas: aritmética módulo 59.

Alfabeto 4: Letras mayúsculas, minúsculas y números: aritmética módulo 69.

Alfabeto 5: Todos los caracteres imprimibles ASCII: aritmética módulo 224.

En los cuatro primeros casos no se tiene en cuenta el carácter del espacio en blanco (valor ASCII 32) puesto que ello entregaría en muchos cifradores clásicos una inapreciable pista al hipotético criptoanalista. Tenga en cuenta que para un texto en castellano en el que el alfabeto considerado sea el de 27 letras más el espacio en blanco, este último presenta una frecuencia de ocurrencia de casi un 20%, siguiéndole muy por detrás las letra *E* y *A*, con valores en el orden del 10%. No obstante, si eliminamos este carácter y ciframos los mensajes solamente con las 27 letras del alfabeto, la letra *E* presenta una frecuencia de aproximadamente un 13%, la letra *A* se alza por encima del 10% y los demás caracteres siguen una distribución característica que será tratada en el apartado siguiente.

En cuanto al quinto alfabeto, hay 224 caracteres imprimibles, desde el valor 32 al 255, contando claro está con el espacio en blanco, o bien 223 sin éste. Si utilizamos como alfabeto de cifrado el código ASCII deberemos tener especial cuidado con los caracteres no imprimibles, caracteres especiales como los de salto de línea, fin de archivo, etc., que luego podrían no ser recuperables. Es decir, la operación de cifrado y descifrado debe considerar estas condiciones de forma que sólo se transmitan los caracteres que pueden imprimirse y no incluir caracteres *extraños* en el criptograma. Evidentemente esto sólo tiene sentido en este tipo de cifradores orientados a caracteres en donde para nada se habla de bits. La cifra moderna es digital y, por tanto, este *problema* no existe. Es más, también podrán *traernos de cabeza* los códigos distintos que usan los sistemas operativos DOS y Windows para caracteres especiales como, por ejemplo, las minúsculas acentuadas o la letra ñ, en tanto el primero utiliza la tabla de códigos ANSI/ASCII y el segundo la tabla ANSI/OEM (véase el Anexo en el libro electrónico http://www.criptored.upm.es/guiateoria/gt_m001a.htm).

intentarse el ataque a partir de las estadísticas del lenguaje. Como veremos en el próximo apartado, en lo que concierne a los cifradores clásicos éstos se dividen en monoalfabéticos y polialfabéticos, en tanto se utilice un único alfabeto para cifrar o más de uno. En tales casos, el análisis de las frecuencias relativas de aparición de los caracteres en el criptograma nos indicará si se trata de uno u otro tipo de cifra.

Aunque los sistemas clásicos estén en desuso, no por ello deben ser pasados por alto por el criptoanalista. En realidad sería bastante poco agradable perder horas de esfuerzo en la intención de romper una cifra, suponiendo de antemano que el criptosistema en cuestión empleado es de los denominados *modernos*, para luego caer en la cuenta que aquel complicado *enigma* se trataba simplemente de un cifrado elemental, que puede romperse fácilmente con herramientas básicas. No quedaríamos muy bien ante nuestros superiores. Por lo tanto, la primera acción que realizará todo criptoanalista será la de *contabilizar* los caracteres que aparecen en el criptograma para obtener información sobre el tipo de cifra, monoalfabético o polialfabético, e intentar aplicar las técnicas que describiremos más adelante para romper dicha cifra. Si esto no entrega los resultados esperados, buscará otros caminos, yendo como es lógico siempre desde la dificultad menor a la mayor.

En la tabla del anexo ya comentado se incluyen las frecuencias relativas de monogramas en el lenguaje castellano módulo 27, esto es considerando sólo las letras mayúsculas. Estos datos nos permiten formar tres grupos de frecuencias relativas: uno de *alta frecuencia*, otro de *frecuencia media* y un tercero de *frecuencia baja*, como se muestra en la Figura 1.8.

| | | | | | |
|---|-------|------------------|------|-----------------|------|
| 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 |
| Frecuencia Alta | | Frecuencia Media | | Frecuencia Baja | |
| Valores de frecuencia relativa expresadas en tanto por ciento | | | | | |

Figura 1.8. Clasificación de frecuencias de caracteres del lenguaje módulo 27.

En la figura anterior, hemos considerado como *Frecuencia Alta* un valor mayor que el 5 por ciento y *Frecuencia Baja* un valor similar o menor que un 1 por ciento. Observe que mezclando las letras de alta frecuencia podemos formar la palabra *ESTIRANDO*. Más adelante volveremos a considerar estos nueve caracteres cuando se aborden las técnicas de criptoanálisis.

Dependiendo del tipo de texto analizado, aparecerán ligeras diferencias, si bien podemos concluir que los valores se mantienen en el rango indicado. Esto quiere decir que es posible considerar, por ejemplo, la letra *L* con más peso que la *D*, incluir en la zona de alta frecuencia la letra *C* en vez de la letra *T*, etc. Piénsese en algún documento que contenga información sobre producción y comercialización de un

determinado bien; es posible que la letra *K* tenga una contribución mayor por el hecho de que aparezca muchas veces la palabra *kilo*; lo mismo en un informe médico de radiología, donde la letra *X* puede tener un mayor peso que el aquí indicado. No obstante, el estudio estadístico de la frecuencia de caracteres tendrá su utilidad sólo en el criptoanálisis de sistemas clásicos por sustitución, en donde supondremos que los mensajes a cifrar se tratará siempre de textos comunes. Es más, en la mayoría de los casos supondremos que tales mensajes contienen solamente caracteres alfabéticos y no del tipo alfanuméricos.

La redundancia del lenguaje no sólo nos dice que existen letras más frecuentes que otras. También nos indica la existencia de digramas comunes, trigramas, poligramas y en general palabras de mayor uso que otras. En la tabla del anexo, se recogen los valores de digramas para un texto en castellano sobre seguridad informática, con más de 40.000 caracteres, el mismo utilizado para la obtención de la tabla de monogramas. En dicha tabla aparecen algunos digramas no existentes en castellano como, por ejemplo, CY, KE y KU. La razón es que el texto utilizado es un documento de 41095 caracteres que habla sobre seguridad informática y virus por lo que aparecen de forma reiterada palabras como *secrecy*, *hackers* y *backup*.

Observando la tabla de digramas, encontramos que los tres digramas con mayor frecuencia relativa en castellano son *DE* (1084), *ES* (1010) y *EN* (901), con cerca del 2,5 por ciento. Asimismo, existirán digramas con frecuencia nula como sería el caso de *QA*, *KK*, *ÑL*, *WZ*, etc., pues no forman parte de palabra alguna ni son término e inicio de dos palabras contiguas. No será éste el caso, por ejemplo, de un digrama como *NF* pues puede ser final e inicio de palabras contiguas en un mensaje como, por ejemplo, *rocínflaco*, *maístínfurioso*, o *atracciónfatal*.

Ejemplo 1.6: *Para el siguiente texto clásico:*

- a) *Encuentre las frecuencias relativas de monogramas.*
- b) *Encuentre los 9 monogramas de mayor frecuencia.*
- c) *Encuentre la frecuencia relativa de digramas.*
- d) *Encuentre los tres digramas más frecuentes.*

"En un lugar de la Mancha, de cuyo nombre no quiero acordarme, no ha mucho tiempo que vivía un hidalgo de los de lanza en astillero, adarga antigua, rocín flaco y galgo corredor. Una olla de algo más vaca que carnero, salpicón las más noches, duelos y quebrantos los sábados, lentejas los viernes, algún palomino de añadidura los domingos, consumían las tres partes de su hacienda. El resto de ella concluían sayo de velarte, calzas de velludo para las fiestas, con sus pantuflos de lo mismo, y los días de entre semana se honraba con su vellorí más fino. Tenía en su casa una ama que pasaba de los cuarenta, y una sobrina que no llegaba a los veinte, y un mozo de campo y plaza, que así ensillaba el rocín como tomaba la podadera. Frisaba la edad de nuestro hidalgo con los cincuenta años; era de complexión recia, seco de carnes, enjuto de rostro, gran madrugador y amigo de la caza. Quieren decir que tenía el sobrenombre de Quijada, o Quesada".

Solución: *Se han contabilizado 730 caracteres.*

- a) *Las frecuencias relativas de monogramas módulo 27 en % para este trozo de texto son:*

| | | | | | |
|---|-------|---|------|---|------|
| A | 14,38 | J | 0,41 | R | 5,75 |
| B | 1,64 | K | 0,00 | S | 7,53 |
| C | 4,38 | L | 6,99 | T | 2,88 |
| D | 5,75 | M | 3,15 | U | 4,93 |
| E | 11,37 | N | 7,53 | V | 1,10 |
| F | 0,68 | Ñ | 0,00 | W | 0,00 |
| G | 1,92 | O | 9,73 | X | 0,14 |
| H | 1,10 | P | 1,51 | Y | 1,23 |
| I | 3,70 | Q | 1,51 | Z | 0,68 |

b) Los nueve monogramas más frecuentes en el texto son:

A, D, E, L, N, O, R, S, U.

c) Los valores absolutos de frecuencia de digramas en el texto se muestran en la siguiente tabla:

| | 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 | 5 | 6 | 7 | 13 | 6 | 1 | 0 | 0 | 0 | 0 | 0 | 13 | 5 | 9 | 0 | 3 | 1 | 5 | 10 | 14 | 0 | 2 | 0 | 0 | 0 | 2 | 2 |
| B | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 4 | 0 | 0 | 1 | 0 | 3 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 |
| D | 10 | 0 | 0 | 1 | 22 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| E | 3 | 1 | 9 | 3 | 2 | 0 | 1 | 1 | 1 | 1 | 0 | 14 | 2 | 18 | 0 | 0 | 1 | 1 | 8 | 11 | 1 | 0 | 3 | 0 | 1 | 1 | 0 |
| F | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| G | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 6 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| H | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| I | 1 | 0 | 1 | 3 | 6 | 0 | 2 | 0 | 0 | 1 | 0 | 2 | 0 | 7 | 0 | 0 | 0 | 0 | 1 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| J | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| K | 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 |
| L | 14 | 0 | 0 | 0 | 4 | 0 | 5 | 0 | 0 | 0 | 0 | 7 | 0 | 0 | 0 | 13 | 1 | 0 | 2 | 1 | 0 | 3 | 0 | 0 | 0 | 0 | 1 |
| M | 6 | 2 | 0 | 0 | 1 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 3 | 3 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| N | 8 | 0 | 4 | 2 | 3 | 1 | 1 | 1 | 0 | 1 | 0 | 4 | 2 | 0 | 0 | 8 | 1 | 0 | 2 | 6 | 8 | 2 | 0 | 0 | 0 | 0 | 1 |
| Ñ | 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 |
| O | 2 | 2 | 5 | 9 | 0 | 0 | 1 | 2 | 0 | 0 | 0 | 2 | 9 | 7 | 0 | 0 | 1 | 3 | 5 | 16 | 3 | 0 | 0 | 0 | 0 | 3 | 1 |
| P | 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 2 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Q | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 0 | 0 | 0 |
| R | 7 | 0 | 0 | 2 | 10 | 0 | 1 | 0 | 2 | 0 | 0 | 0 | 2 | 3 | 0 | 8 | 0 | 1 | 1 | 0 | 2 | 2 | 0 | 0 | 0 | 1 | 0 |
| S | 7 | 1 | 4 | 8 | 6 | 2 | 0 | 0 | 1 | 0 | 0 | 3 | 2 | 1 | 0 | 2 | 2 | 0 | 0 | 1 | 6 | 5 | 3 | 0 | 0 | 1 | 0 |
| T | 3 | 0 | 0 | 0 | 6 | 0 | 0 | 0 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 0 | 0 | 4 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| U | 3 | 0 | 2 | 1 | 11 | 1 | 2 | 1 | 3 | 0 | 0 | 0 | 1 | 6 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| V | 2 | 0 | 0 | 0 | 4 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| W | 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 |
| X | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Y | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 2 | 1 | 1 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| Z | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

d) Los tres digramas más frecuentes del texto son DE con 22 apariciones, EN con 18 y OS que aparece 16 veces.

En el ejemplo anterior, a pesar de que el texto no tiene la longitud que sería recomendable para obtener unos resultados que sean fieles a la realidad de la *ratio* y *redundancia* del lenguaje, sí deja entrever una tendencia marcada del mayor peso de algunas letras y conjunto de letras. De las 9 letras de mayor peso en este texto, 7 corresponden a la clasificación de *Alta Frecuencia* que habíamos hecho.

En cuanto a los digramas, existe una mayor dispersión como es natural porque el texto analizado es muy corto. No obstante, para este texto con 726 digramas, dos de los tres digramas más comunes del texto, DE con un 3,0 % y EN con el 2,5 %, son

también los más frecuentes en el lenguaje castellano con valores de frecuencia muy similares.

1.3. CLASIFICACIÓN DE LOS CRIPTOSISTEMAS CLÁSICOS

El cifrado o cifra es una técnica para ocultar un mensaje y evitar que sea legible si éste es interceptado por una persona no autorizada. Por lo tanto, el objetivo básico es mantener seguros unos datos dentro de un entorno como puede ser una línea de transmisión o un sistema de almacenamiento que ya hemos visto es inseguro. Como protección utilizaremos métodos o algoritmos para cifrar la información. En una primera aproximación, en este caso bajo el punto de vista histórico, clasificaremos estos métodos de cifra como *Criptosistemas Clásicos* y *Criptosistemas Modernos*.

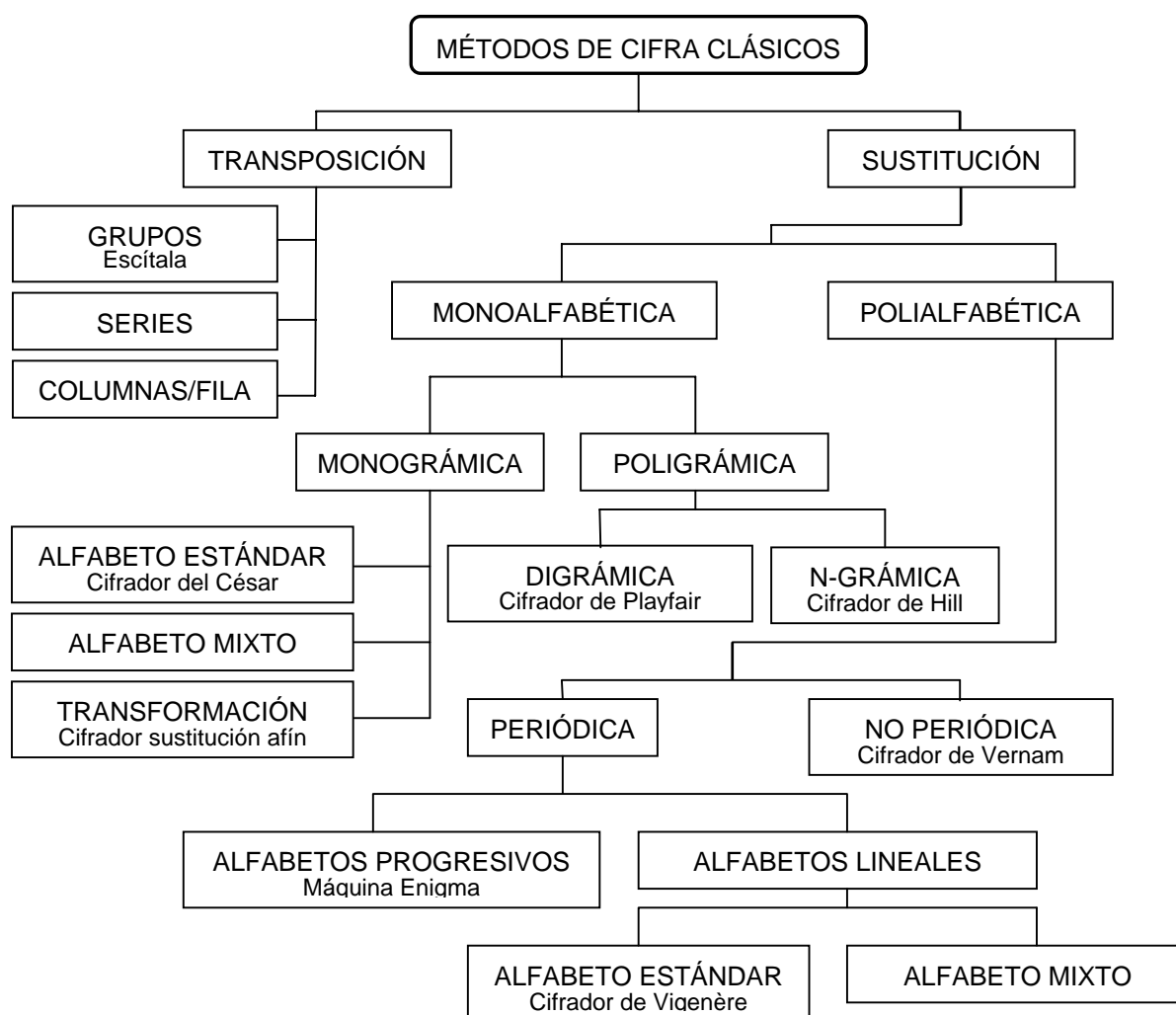


Figura 1.9. Clasificación de los métodos clásicos de cifra y algunos ejemplos.

Los métodos clásicos son aquellos en los que, además de las máquinas dedicadas para cifrar como las estudiadas en el apartado 1.1, se usan por separado

técnicas de sustitución y transposición aplicadas a los caracteres del texto en claro. Las técnicas criptográficas utilizadas en este caso son en su totalidad orientadas a sistemas de clave secreta, generalmente manteniendo también en secreto el algoritmo, incluso en el caso en que el cifrador cuente con una clave secreta. La operación de cifra se realiza sobre caracteres alfanuméricos, por lo general alfabéticos, y en ese mismo formato se transmiten o almacenan.

La Figura 1.9 muestra una clasificación de los sistemas de cifra clásicos, en donde se incluyen algunos cifradores típicos a modo de ejemplo. Estos sistemas de cifra se clasificarán, básicamente, en aquellos que utilizan técnicas de *sustitución* y aquellos que utilizan técnicas de *transposición* sobre los caracteres de un texto en claro, ambas técnicas propuestas por *Shannon* para lograr la *confusión* y *difusión*, respectivamente.

NOTA: *En lo que sigue del texto, para todos los ejemplos supondremos que el texto cifrado viene agrupado en una cantidad fija de caracteres y que tanto los espacios en blanco como los signos de puntuación no serán tomados en cuenta en el proceso de cifra. La agrupación por defecto de los elementos del criptograma será de cinco caracteres.*

Los cifradores por *transposición* utilizan la técnica de *permutación* de forma que los caracteres del texto se reordenan mediante un algoritmo específico. Un caso representativo de esta transformación -que será analizado más detenidamente en un apartado próximo- sería transmitir el mensaje en bloques de cinco caracteres pero reordenados (permutados) éstos de forma que su posición en el criptograma sea, por ejemplo, 43521; es decir, el cuarto carácter del bloque en claro se transmite primero, a continuación el tercero, después el quinto, luego el segundo y, por último, el primero. Esta operación se repetirá en cada bloque de 5 caracteres del mensaje. Por lo tanto, la transposición implica que los caracteres del criptograma serán exactamente los mismos que los del texto en claro.

Ejemplo 1.7: *Cifre mediante transposición de bloques de cinco caracteres el siguiente mensaje, usando la permutación 43521.*
M = AL GRITO DE VIVA ZAPATA SE ARMÓ UNA GORDA.

Solución: *Siguiendo la permutación indicada, se obtiene:*
M = ALGRI TODEV IVAZA PATAS EARMO UNAGO RDAXX
C = RGILA EDVOT ZAAVI ATSAP MROAE GAONU XAXDA

Los cifradores por *sustitución* utilizan la técnica de *modificación* de cada carácter del texto en claro por otro correspondiente al alfabeto de cifrado. Si el alfabeto de cifrado es el mismo que el del mensaje o bien único, hablamos entonces de cifradores *monoalfabéticos*; es decir, existe un único alfabeto en la operación de transformación del mensaje en criptograma. Por el contrario, si en dicha operación intervienen más de un alfabeto, se dice que el cifrador es *polialfabético*.

¿Cómo es posible utilizar más de un alfabeto en la operación de cifrado? La respuesta es muy sencilla y la abordaremos a continuación. En el cifrador del César, por ejemplo, la letra *A* del texto en claro se cifraba *siempre* como la letra *D*; es por tanto

un cifrador monoalfabético. A continuación desarrollaremos un algoritmo sencillo para usar más de un alfabeto. Suponga que deseamos diseñar un algoritmo de cifrado similar al del César, de forma que a los caracteres impares aplicamos un desplazamiento de 15 espacios a la derecha del alfabeto y a los caracteres pares 10 espacios también a la derecha según el ejemplo 1.8.

Ejemplo 1.8: *Utilizando el algoritmo propuesto en el párrafo anterior, se pide cifrar el mensaje $M = \underline{DISFRUTAN} \underline{VACACIONES} \underline{EN} \underline{EL} \underline{MES} \underline{DE} \underline{AGOSTO}$.*

Solución: *Se cifrará de la siguiente forma: los caracteres en posiciones impares subrayados (dsrtn...aot) se desplazan 15 lugares y los caracteres en posiciones pares (ifuav...gso) se desplazan 10 lugares. Usando entonces las congruencias $m_i + 15 \bmod 27$ para los primeros y $m_i + 10 \bmod 27$ para los segundos, se obtiene:*

$C = RRHOG \ EIKBF \ OMOMW \ YBÑHÑ \ BÑZVS \ CRÑOP \ DCIY$.

En el criptograma anterior ha dejado de existir una correspondencia única entre los caracteres del texto en claro y los de un alfabeto de cifrado: la letra *A* se cifra como el carácter *K* o como el carácter *O*, dependiendo se encuentre en el texto en una posición par o impar, respectivamente. Otro tanto ocurre para las letras *I*, *S*, *E*, y *O*. Por otra parte, el criptograma comienza con el digrama *RR* que corresponde a dos letras distintas del texto en claro.

Al aplicar dos desplazamientos diferentes, hemos utilizado dos alfabetos de cifrado distintos, de ahí que algunas letras según su posición se cifren como dos letras distintas. En la Figura 1.10 se muestran los dos alfabetos utilizados en el ejemplo; A_1 para los caracteres impares y A_2 para los pares.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 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_i | 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_1 | 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 |
| A_2 | 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 |

Figura 1.10. Alfabetos utilizados en el ejemplo.

Esta forma de cifrar dará lugar, entre otros, al cifrador de Vigenère que veremos más adelante. Si este fuera el caso, la clave utilizada habría sido $K = OK$, puesto que el equivalente numérico de la letra *O* es 15 y el de la letra *K* es igual a 10, los valores de desplazamiento utilizados para caracteres impares y pares, respectivamente. Observe que al sumarse la clave al texto y como el equivalente de la letra *A* es igual a cero, se cumplen las congruencias $A+O = O$ y $A+K = K$, los dos valores con los que comienzan los alfabetos en cuestión.

La sustitución polialfabética puede ser *periódica* como en el caso del ejemplo anterior cuyo período es dos (el tamaño de la clave) o bien *no periódica*, cuando la clave en cuestión es tan larga como el mensaje.

Tanto en el caso monoalfabético como en el polialfabético, se realiza la transformación $E_K(M)$ sobre cada uno de los caracteres del texto en claro de forma

independiente; es decir, la operación se realiza carácter a carácter o lo que es lo mismo a través de *monogramas*. También es posible cifrar un texto en claro utilizando bloques de más de un carácter. En este caso se hablará de cifradores *digrámicos* que cifran el texto cada dos caracteres, *trigrámicos* en bloques de tres caracteres y, en general, *poligrámicos*. Al hablar de *alfabetos mixtos*, nos referiremos al uso de alfabetos que contienen caracteres distintos al propio del lenguaje, por ejemplo mediante el uso de símbolos. La Figura 1.11 muestra un posible alfabeto de cifrado mixto.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|----------------|---|---|---|---|---|---|---|----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M _i | 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 |
| Ci | < | > | (|) | & | % | / | \$ | ¿ | T | D | A | C | U | E | N | ? | ♣ | ♦ | ♥ | ♠ | @ | # | [| α | β |] |

Figura 1.11. Ejemplo de alfabeto de cifrado mixto.

1.4. CIFRADORES POR SUSTITUCIÓN MONOGRÁMICA MONOALFABETO

1.4.1. Cifradores por sustitución

Un cifrador por sustitución es aquel que sustituye cada carácter del texto en claro por otro carácter en el texto cifrado o criptograma. Esta es la forma de aplicar el principio de *confusión* propuesto por Shannon en cuanto que oculta el texto en claro a cualquier intruso mediante sustituciones, excepto para el destinatario, que conoce el algoritmo y la clave que le permite descifrar el criptograma para recuperar el mensaje. Los cifradores por sustitución se pueden clasificar en tres grupos; a saber, *sustitución monográfica monoalfabeto*, *sustitución monográfica polialfabeto* y *sustitución poligrámica*.

En los cifradores por *sustitución monográfica monoalfabeto*, el cifrado se realiza mediante un algoritmo que hace corresponder una letra del texto en claro a una única letra del criptograma, es decir, cifra monogramas. De ahí su denominación de cifrador monográfico. En cuanto al término monoalfabeto, quiere decir que se utiliza un único alfabeto de cifrado, el mismo que el del texto en claro o uno mixto, pero distribuido bien de forma aleatoria o bien a través de una transformación matemática. Luego, si a la letra *M* del texto en claro le corresponde la letra *V* o el símbolo *#* del alfabeto de cifrado, se cifrará siempre igual pues existe una única equivalencia o, lo que es lo mismo, un único alfabeto de cifrado.

Por su parte, en los cifradores por *sustitución monográfica polialfabeto*, la operación de cifra también se realiza carácter a carácter, es decir por monogramas. No obstante, como ya hemos mencionado en apartados anteriores, a través de una clave, algoritmo o mecanismo, se obtienen varios alfabetos de cifrado de forma que una misma letra puede cifrarse con caracteres distintos, dependiendo de su posición dentro del texto en claro.

Por último, los cifradores por *sustitución poligrámica* tratan el mensaje en bloques de dos o más caracteres sobre los que se aplica la transformación del criptosistema en cuestión, sustituyendo ngramas del mensaje por ngramas de texto

cifrado.

1.4.2. El cifrador del César

Tal vez el cifrador monoalfabético por sustitución más famoso es el denominado *Cifrador del César*, uno de los cifradores más antiguos, atribuido al emperador romano *Julio César*. Se trata de un criptosistema en el que se aplica un desplazamiento constante igual de b caracteres sobre el texto en claro, obteniéndose así el criptograma buscado. Este tipo de cifradores, llamados genéricamente *cifradores monoalfabéticos por desplazamiento puro* o adición, toman en el caso del cifrador del César un valor de desplazamiento b igual a 3. Por lo tanto, este cifrador del César tendrá el alfabeto de cifrado ya representado en la Figura 1.3 con su equivalente numérico, es decir:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 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_i | 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_i | 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 |

Figura 1.12. Equivalente numérico del alfabeto de cifrado del César.

Ejemplo 1.9: Con la tabla del cifrador del César de la Figura 1.12, cifre el mensaje:
 $M = \text{CÉSAR EL EMPERADOR HA SIDO ASESINADO.}$

Solución: Aplicando a cada carácter M_i su equivalente C_i , se obtiene:
 $C = \text{FHVDU HÑHOS HUDGR UKDVL GRDVH VLPDG R.}$

Este sistema de cifra sencillo, apropiado e incluso bastante ingenioso para la época, presenta un nivel de seguridad muy débil; de hecho su distancia de unicidad es muy baja. En el apartado siguiente encontraremos este valor y se demostrará que el criptoanálisis de este cifrador es verdaderamente elemental, un pasatiempo.

• Cifrador del César con clave

Para aumentar la seguridad o, lo que es lo mismo, la distancia de unicidad de este cifrador, podemos incluir en el alfabeto de cifrado una clave de la siguiente forma: la clave K consiste en una palabra o frase que se escribe a partir de una posición p_0 del alfabeto en claro. Si la clave es $K = \text{ESTOY ABURRIDO}$ y la posición inicial $p_0 = 3$ tenemos:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| | 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_i | 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 | | | | E | S | T | O | Y | A | B | U | R | I | D | | | | | | | | | | | | | |

Los caracteres repetidos de la clave no se escriben. Una vez escrita ésta en la posición indicada, se añaden las demás letras del alfabeto en orden y de forma modular, para obtener así el alfabeto de cifrado, como se observa.

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

Figura 1.13. Ejemplo de alfabeto de cifrador del César con clave.

En este tipo de cifrado se deja de cumplir la condición de desplazamiento constante, característica en el sistema del César primario.

Ejemplo 1.10: Con el cifrador del César con clave cifre el siguiente mensaje usando como clave $K = \text{POBRE CHUCHO SIBERIANO}$ con $p_0 = 2$.
 $M = \text{A PERRO FLACO TODO SON PULGAS.}$

Solución: El alfabeto de cifrado resultante será:
 M_i 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_i Y Z P O B R E C H U S I A N D F G J K L M Ñ Q T V W X
 Luego, $C = \text{YGBKK FRIYP FMFOF LFNGÑ IEYL.}$

Como es de esperar, al tener un mayor número de combinaciones de alfabetos, existe una mayor incertidumbre respecto de la clave. La distancia de unicidad de este cifrador será mayor y, por consiguiente, el sistema presentará una mayor fortaleza.

• Cifradores tipo César con alfabetos mixtos

En este tipo de cifrado se aplica también una relación única entre un elemento del alfabeto en claro y un elemento del alfabeto de cifrado, salvo que este último puede contener otros caracteres o símbolos que no pertenezcan al alfabeto del mensaje.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M_i | 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_i | [| < | > | { | } | x | ♥ | ♦ | ♣ | ♠ | ≠ | # | @ | % | & | (|) | = | > | < | 0 | 1 | 2 | 3 | 4 | 5 |] |

Figura 1.14. Cifrador del César con alfabeto mixto.

Ejemplo 1.11: Cifre con el alfabeto mixto de la Figura 1.14 el siguiente mensaje:
 $M = \text{EN EL ESCARABAJO DE ORO APARECEN SIGNOS DISTINTOS A LOS DEL TEXTO EN CLARO.}$

Solución: Siguiendo el alfabeto de cifra indicado se obtiene el criptograma:
 $C = \text{\%}\#\} <\>[\llbracket \spadesuit (\{ \} > ([\] > \} \} \% < \clubsuit \heartsuit \% (< \clubsuit < 0 \clubsuit \% 0 (< \# (< \{ \} \# 0 \} 4 0 \{ \} \% \{ \# [> (.}$

Uno de los casos más interesantes de este tipo de cifradores con alfabetos distintos desde el punto de vista histórico lo encontramos en los grabados realizados en 1794 en una lápida del cementerio de *Trinity*, un distrito de Nueva York. El mensaje consistía en un conjunto de símbolos como se indica a continuación.

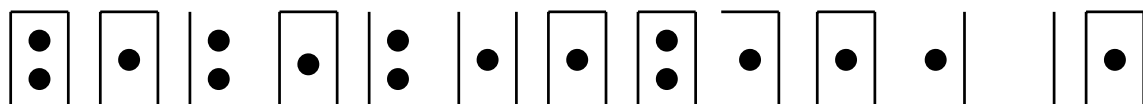


Figura 1.15. Criptograma de la lápida del cementerio de Trinity.

Sólo 100 años después, en 1896, se logra descriptar este enigma aplicando nociones básicas de estadística y redundancia del lenguaje. Esto indica que, por lo

menos para aquella época, el criptosistema empleado aunque hoy en día muy inocente, cumplió con creces el principio de/ la caducidad de la información. La resolución del criptograma, de gran dificultad debido al pequeño tamaño del mismo, sigue la siguiente clave de asignación de figuras geométricas a los caracteres:

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

Figura 1.16. Clave para la solución del criptograma de Trinity.

Siguiendo entonces las claves de la Figura 1.16, encontramos que el mensaje del criptograma en cuestión es $M = \text{REMEMBER DEATH}$. La dificultad para llegar a este resultado es obvia pues se cuenta con un criptograma demasiado pequeño. En su contra está el hecho de que, precisamente la letra *E*, la más significativa del inglés, aparece cuatro veces y que el texto que se ha cifrado tiene mucho que ver con el epitafio que alguna *mente inquieta y algo retorcida* pondría en su tumba.

Un sistema muy similar al indicado en la Figura 1.16, denominado *Freemasons*, es utilizado por las logias de la masonería. El lector interesado en este tipo de cifrados históricos puede consultar el libro *Cryptology: Machines, History & Methods*, cuya referencia ya ha sido indicada en este capítulo.

Este tipo de cifrado, -precisamente el utilizado por *Edgar Allan Poe* en su famoso cuento *El escarabajo de oro*- no por ser más *espectacular* en la presentación del criptograma es más seguro que el del César con clave: en el fondo, son exactamente lo mismo. Como veremos en el próximo apartado de criptoanálisis, con unas cuantas decenas de caracteres de criptograma, romper este tipo de cifra puede ser algo realmente fácil.

1.4.3. Criptoanálisis del cifrado del César

Comentábamos en el apartado anterior que la distancia de unicidad del cifrador del César con clave será mayor que en el caso sin clave y, por consiguiente, también su seguridad. No obstante, al producirse en ambos casos una sustitución fija de cada carácter del alfabeto en claro por un único carácter del alfabeto de cifrado, el criptograma podrá romperse fácilmente aplicando técnicas de estadística del lenguaje, siempre y cuando contemos con una cantidad suficiente de texto cifrado. En un capítulo anterior vimos que la distancia de unicidad venía dada por la relación entre la entropía de la clave $H(K)$ y la redundancia del lenguaje D . Volveremos a encontrar este valor para cifradores del tipo César.

Ejemplo 1.12: *Encuentre la distancia de unicidad de un cifrador del César.*

Solución: *Si $n = 27$, existirán sólo 26 posibles combinaciones de alfabetos, por lo tanto $H(K) = \log_2 26 = 4,70$. Como la redundancia D era igual a 3,4*

entonces se tiene que $N = H(K)/D \approx 4,70/3,4 \approx 1,38$. Por lo tanto, necesitamos como mínimo dos caracteres.

Para cifrados del César sin clave, una forma elemental de criptoanálisis consiste en escribir bajo el texto cifrado todas las combinaciones de frases, con o sin sentido, que se obtienen al aplicar a dicho criptograma desplazamientos de 1, 2, 3, ..., n-1 caracteres, siendo n el número de caracteres del alfabeto utilizado. Una de estas combinaciones dará con el texto en claro y esto será válido independientemente del valor asignado a la constante de desplazamiento. Retomemos el Ejemplo 1.9 en el que teníamos el siguiente par mensaje/criptograma, respetando los espacios en blanco para una mayor claridad:

M = CESAR EL EMPERADOR HA SIDO ASESINADO.
C = FHVDU HÑ HOSHUDGRU KD VLGR DVHVLPDGR.

Como nuestro alfabeto contiene 27 caracteres tenemos el cuadro de posibles mensajes a partir del criptograma que se muestra en la Figura 1.17. En esta figura, la única solución con sentido corresponde a un desplazamiento de 24 caracteres en la operación de descifrado. Si al cifrar el mensaje M nos hemos desplazado 3 espacios hacia delante para obtener el criptograma C, representado en la posición b = 0 de la figura, para descifrarlo habrá que desplazarse 3 caracteres hacia atrás o bien, de acuerdo con la modularidad, un desplazamiento hacia delante de $27-3 = 24$ espacios.

| | | | |
|----|--------|--|-----------------------|
| C: | b = 0 | FHVDU HÑ HOSHUDGRU KD VLGR DVHVLPDGR | <u>Texto cifrado</u> |
| | b = 1 | GIWEV IO IPTIVEHSV LE WMHS EWIWMQEHS | |
| | b = 2 | HJXFW JP JQUJWFITW MF XNIT FXJXNRFIT | |
| | b = 3 | IKYGX KQ KRVKXGJUX NG YÑJU GYKYÑSGJU | |
| | b = 4 | JLZHY LR LSWLYHKVY ÑH ZOKV HZLZOTHKV | |
| | b = 5 | KMAIZ MS MTXMZILWZ OI APLW IAMAPUILW | |
| | b = 6 | LNBJA NT NUYN AJMXA PJ BQMX JBNBQVJMX | |
| | b = 7 | MÑCKB ÑU ÑVZÑBK NYB QK CRNY KCÑCRWK NY | |
| | b = 8 | NODLC OV OWAOC LÑZC RL DSÑZ LDODSXLÑZ | |
| | b = 9 | ÑPEMD PW PXBPDMOAD SM ETOA MEPETYMOA | |
| | b = 10 | OQFNE QX QYCQENPBE TN FUPB NFQFUZNPB | |
| | b = 11 | PRGÑF RY RZDRFÑQCF UÑ GVQC ÑGRGVAÑQC | |
| | b = 12 | QSHOG SZ SAESGORDG VO HWRD OHSWBORD | |
| | b = 13 | RTIPH TA TBFTHPSEH WP IXSE PITIXCPSE | |
| | b = 14 | SUJQI UB UCGUIQTFI XQ JYTF QJUJYDQTF | |
| | b = 15 | TVKRJ VC VDHVJRUGJ YR KZUG RKVKZERUG | |
| | b = 16 | UWLSK WD WEIWKS VHK ZS LAVH SLWLAFSVH | |
| | b = 17 | VXMTL XE XFJXL TWIL AT MBWI TMXMBGTWI | |
| | b = 18 | WYNUM YF YGKYMUXJM BU NCXJ UNYNCHUXJ | |
| | b = 19 | XZÑVN ZG ZHLZNVYKN CV ÑDYK VÑZÑDIVYK | |
| | b = 20 | YAOWÑ AH AIMAÑWZLÑ DW OEZL WAOEJWZL | |
| | b = 21 | ZBPXO BI BJNBXAMO EX PFAM XPBPFKXAM | |
| | b = 22 | ACQYP CJ CKÑCPYBNP FY QGBN YQCQGLYBN | |
| | b = 23 | BDRZQ DK DLODQZCÑQ GZ RHCÑ ZRDRHMZCÑ | |
| M: | b = 24 | CESAR EL EMPERADOR HA SIDO ASESINADO | <u>Texto en claro</u> |

| | |
|--------|--------------------------------------|
| b = 25 | DFTBS FM FNQFSBEPs IB TJEP BTFTJÑBEP |
| b = 26 | EGUCT GN GÑRGTCFQT JC UKFQ CUGUKOCFQ |

Figura 1.17. Criptoanálisis del cifrador del César.

Para los cifradores por desplazamiento puro como el del César, se cumplirá por tanto la siguiente operación de descifrado (D) a partir de un cifrado (E) en el anillo n:

$$D_b = E_{n-b} \Rightarrow D_b = E_{27-b} \quad \boxed{1.4}$$

De lo visto anteriormente, es fácil deducir que un sistema de cifra por sustitución monoalfabética como el del César presenta un nivel de seguridad mínimo en tanto que para romperlo nos ha bastado con un *lápiz*, *papel* y un poco de *paciencia* para confeccionar el cuadro anterior, nada del otro mundo como puede ver. Esta debilidad se debe a que el número de desplazamientos posibles es muy pequeño al contar sólo con los 26 valores que corresponden a los caracteres del alfabeto; esto es, se cumple que $1 \leq b \leq 26$, pues un desplazamiento igual a cero o bien múltiplo de veintisiete sería igual que transmitir en claro.

En el caso en que el cifrador del César tenga una clave, el ataque anterior no proporciona ninguna solución porque el desplazamiento deja de ser constante y, por lo tanto, es imposible establecer una relación matemática única y directa entre el alfabeto en claro y el alfabeto de cifrado. El único camino que nos queda consiste en llevar las estadísticas del lenguaje al criptograma, observando por ejemplo la frecuencia relativa de aparición de los caracteres en el texto cifrado. Al contrario del método anterior, válido solamente para desplazamientos puros sin clave, este tipo de ataque estadístico será válido tanto para los cifrados de tipo monoalfabético con clave como para aquellos que no la tienen. Ahora bien, en la gran mayoría de los casos será necesario contar con una cantidad de criptograma bastante mayor que la del ejemplo anterior y, cómo no, una *pizca* de intuición y un poco de suerte.

Ejemplo 1.13: *Calcule la distancia de unicidad de un cifrador del César con clave.*

Solución: *Si el alfabeto tiene n caracteres, existirán n! combinaciones posibles de n elementos, luego $N = H(K)/D = (\log_2 n!)/D$. Utilizando la aproximación de Sterling, $\log_2 n! \approx n \log_2(n/e)$, la distancia de unicidad será $N \approx n \log_2(n/e)/D$. Para $n = 27$ se tiene: $N \approx 27 \log_2(27/e)/3,4 \approx 27,4$ caracteres.*

Para el cifrador del César, al establecerse en la operación de cifrado una correspondencia directa entre los caracteres del texto en claro y del alfabeto de cifrado, se mantiene la misma relación de frecuencia relativa característica del lenguaje. Por lo tanto, es muy probable que la letra C_i del texto cifrado con mayor frecuencia relativa se corresponda con la letra M_i de mayor frecuencia relativa del lenguaje. Esto es, si la letra W es la de mayor frecuencia en el criptograma, podemos suponer con muy buenas expectativas de éxito que sea la letra E del texto en claro y que, por lo tanto, el desplazamiento aplicado haya sido igual a 19, la distancia que separa ambas letras en el alfabeto. Como ya hemos comentado, estas suposiciones sólo tendrán cierta validez si la cantidad de texto cifrado es grande y por tanto se cumplen las propiedades estadísticas del lenguaje. En el fondo se está realizando una comparación de la

distribución de frecuencias de todos los elementos del criptograma con la característica del lenguaje, con el objeto de encontrar ese desplazamiento constante.

Otra forma de atacar un cifrado por desplazamiento puro con o sin clave es buscar digramas, trigramas, ngramas o poligramas y en general palabras características del lenguaje, para asociar un conjunto de caracteres del criptograma con otro conjunto de caracteres del texto en claro. Obviamente este criptoanálisis es también válido para atacar cifrados con alfabetos mixtos como se observa en el siguiente ejemplo.

Ejemplo 1.14: *Describe el siguiente criptograma cifrado con un alfabeto mixto:*

$C = \spadesuit >] \clubsuit 0 \heartsuit 34 \spadesuit 3 44] > \} > \{ : \odot \spadesuit \{ 4 \heartsuit 3 > 34] 02 014 : 0 34 (\spadesuit 4 \heartsuit \clubsuit > \spadesuit) 0 \{ \} 4 03401 4 \} 034 \spadesuit 344] 0 \heartsuit 1 >]$.

Solución: *Encontramos las siguientes frecuencias relativas en los caracteres:*

| | | | | |
|------------------|-------------|------------------------|----------------|-------|
| 4 = 14 | 0 = 9 | 3 = 8 | > = 6 |] = 6 |
| \heartsuit = 4 | { = 3 | \spadesuit = 3 | \diamond = 3 | 1 = 3 |
| \clubsuit = 2 |) = 2 | : = 2 | } = 2 | (= 1 |
| 2 = 1 | \odot = 1 | Total : 70 caracteres. | | |

Suponiendo que los caracteres 4 y 0, de mayor frecuencia relativa en el criptograma, se corresponden con las letras E y A del alfabeto, se obtiene el siguiente criptograma parcial:

$M_1 = \spadesuit >] \clubsuit A \heartsuit 3E \spadesuit 3 EE] > \} > \{ : \odot \spadesuit \{ E \heartsuit 3 > 3E] A2 A1E : A 3E (\spadesuit E \heartsuit \clubsuit > \spadesuit) A \{ \} E A3EA1 E \} A3E \spadesuit 3EE] A \heartsuit 1 >]$.

Si por alguna pista sospechamos que]A2A1E:A sea LACABEZA, tenemos otros 4 caracteres y el criptograma sería en una segunda aproximación:

$M_2 = \spadesuit > L \clubsuit A \heartsuit 3E \spadesuit 3 EE] > \} > \{ : \odot \spadesuit \{ E \heartsuit 3 > 3ELAC ABEZA 3E (\spadesuit E \heartsuit \clubsuit > \spadesuit) AL \{ \} E A3EAB E \} A3E \spadesuit 3EEL A \heartsuit B > L$.

Ya tenemos entonces 6 caracteres con posible equivalencia:

0 = A; 1 = B; 2 = C; 4 = E;] = L y : = Z.

Del alfabeto anterior podríamos suponer que 3 = D. Si además pensamos que la cadena ELA \heartsuit B>L es ELARBOL, el resultado son tres correspondencias nuevas: 3 = D; \heartsuit = R y > = O. Obtenemos el alfabeto de cifrado que se indica y una tercera aproximación del criptograma:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|--------------|---|---|---|---|---|---|---|---|
| M_i | 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_i | 0 | 1 | 2 | 3 | 4 | _ | _ | _ | _ | _ | _ |] | _ | _ | _ | > | _ | _ | \heartsuit | _ | _ | _ | _ | _ | _ | _ | : |

$M_3 = \spadesuit OL \clubsuit A RDE \spadesuit D EEL > \} > \{ Z \odot \spadesuit \{ ERDO DELAC ABEZA DE (\spadesuit E R \clubsuit O \spadesuit) AL \{ \} E ADEAB E \} ADE \spadesuit DEEL ARBOL$.

Si ABE}ADE \spadesuit DEELARBOL es ABEJADESDEELARBOL, obtenemos dos nuevas correspondencias entre caracteres } = J y \diamond = S y el siguiente criptograma:

$M_4 = SOL \clubsuit A RDESD EEL > J > \{ Z \odot \spadesuit \{ ERDO DELAC ABEZA DE (\spadesuit E R \clubsuit O \spadesuit) AL \{ \} E ADEAB E \} ADE SDEEL ARBOL$.

La resolución final del criptograma a estas alturas parece ya algo trivial. El alfabeto utilizado en este ejemplo y el mensaje son los que se indican:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---------|--------------|------------|-------------|--------------|----|---|---|---|---|
| M_i | 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_i | 0 | 1 | 2 | 3 | 4 | 5 | « | » | { | } | [|] | (|) | < | > | J | \odot | \heartsuit | \diamond | \clubsuit | \spadesuit | \$ | & | @ | # | : |

$M =$ SOLTAR DESDE EL OJO IZQUIERDO DE LA CABEZA DE MUERTO UNA LÍNEA DE ABEJA DESDE EL ÁRBOL.

La técnica de criptoanálisis del ejemplo anterior es precisamente la usada por *Allan Poe* en su cuento "*El escarabajo de oro*". El texto corresponde a una parte de ese *enigmático mensaje* con un alfabeto de cifrado ligeramente distinto.

Como conclusión podemos afirmar que, incluso incluyendo una clave en la cifra, por complicada y larga que ésta sea, estos criptosistemas monoalfabéticos son muy vulnerables a los ataques de un criptoanalista. Si además se cuenta con la ayuda de un simple ordenador, el ataque y posterior solución a estos criptogramas se convierte en la práctica en un divertido juego.

1.4.4. Cifradores genéricos por sustitución

Ya hemos comentado en el apartado anterior qué se entiende por un cifrador monoalfabético por sustitución, por lo menos para el caso del cifrador del César en el que la sustitución de los caracteres se obtiene por medio de un desplazamiento constante en el alfabeto.

A continuación analizaremos los cifradores monoalfabéticos genéricos, también conocidos como cifradores de transformaciones afines. En este caso la operación de sustitución de los caracteres del alfabeto puede obtenerse de forma matemática aplicando la siguiente expresión de equivalencia:

$$C_i = (a * M_i + b) \bmod n \quad \boxed{1.5}$$

en donde a se conoce como la *constante de decimación*, b *constante de desplazamiento* y n es el *orden del grupo*. Observe que de acuerdo con la ecuación (1.5) la relación matemática del cifrador del César será: $C_i = (M_i + 3) \bmod n$.

¿Puede utilizarse cualquier valor de a y b en la ecuación anterior? La respuesta es no. En primer lugar, es obvio que a no puede ser igual a cero pues no existiría una equivalencia de alfabetos por lo que deberá cumplirse que $a \geq 1$. Por otra parte, para la existencia de inversos deberá cumplirse que los valores de la constante a y el módulo n sean primos entre sí; es decir $\text{mcd}(a, n) = 1$. Al trabajar en módulo $27 = 3^3$, los valores permitidos de la constante de decimación a serán los 18 elementos del CRR(27) que no tengan como factor común el número 3, es decir: 1, 2, 4, 5, 7, 8, 10, 11, 13, 14, 16, 17, 19, 20, 22, 23, 25 y 26.

En cuanto a la constante de desplazamiento b , ésta puede tomar cualquier valor comprendido entre 0 y $n-1$ pues se asegura en todo momento la existencia del inverso para la adición. Desplazamientos mayores que $n-1$ caerán dentro del mismo anillo por lo que su valor se reduce módulo n . Por ejemplo, un desplazamiento de $b = 32$ espacios equivale a $(32 - k * n) = (32 - 1 * 27) = 5$ espacios efectivos. Así mismo, un desplazamiento negativo (caracteres hacia la izquierda del alfabeto) puede trasladarse a su equivalente en el intervalo $[0, n-1]$. Por lo tanto, podemos decir que un desplazamiento de $b = -8$ caracteres equivale a $(-8 + k * n) = (-8 + 1 * 27) = 19$ caracteres hacia la derecha.

De esta manera, diremos que cuando la constante de decimación a es igual a la unidad, el cifrador genérico de sustitución se transforma en uno de *desplazamiento puro*; cuando la constante de desplazamiento b es igual a cero hablamos de cifradores por *decimación pura* y cuando se cumple que la constante a es mayor que la unidad y b es distinto de cero, la cifra es por *sustitución afín*.

• Cifradores por desplazamiento puro

Corresponden a los denominados cifradores tipo César ya vistos en el apartado anterior por lo que no vamos a repetir lo allí comentado. Las operaciones de cifra y descifrado serán:

$$C_i = (M_i + b) \bmod n \quad 1.6$$

$$M_i = (C_i - b) \bmod n \quad 1.7$$

$$M_i = (C_i + n - b) \bmod n \quad 1.8$$

Las ecuaciones (1.7) y (1.8) son equivalentes en el cuerpo n .

Ejemplo 1.15: a) Con $n = 27$ y un desplazamiento $b = 15$, cifre el mensaje $M = \text{SALVE CÉSAR}$ utilizando la ecuación (1.6). b) Descifre el criptograma mediante la ecuación (1.7).

Solución:

a) $C_i = (m_i + 15) \bmod 27$

$$C_{01} = (S+15) \bmod 27 = (19+15) \bmod 27 = 34 \bmod 27 = 07 = H$$

$$C_{02} = (A+15) \bmod 27 = (00+15) \bmod 27 = 15 \bmod 27 = 15 = O$$

$$C_{03} = (L+15) \bmod 27 = (11+15) \bmod 27 = 26 \bmod 27 = 26 = X$$

$$C_{04} = (V+15) \bmod 27 = (21+15) \bmod 27 = 36 \bmod 27 = 09 = J$$

$$C_{05} = (E+15) \bmod 27 = (04+15) \bmod 27 = 19 \bmod 27 = 19 = S$$

$$C_{06} = (C+15) \bmod 27 = (02+15) \bmod 27 = 17 \bmod 27 = 17 = Q$$

$$C_{07} = (E+15) \bmod 27 = (04+15) \bmod 27 = 19 \bmod 27 = 19 = S$$

$$C_{08} = (S+15) \bmod 27 = (19+15) \bmod 27 = 34 \bmod 27 = 07 = H$$

$$C_{09} = (A+15) \bmod 27 = (00+15) \bmod 27 = 15 \bmod 27 = 15 = O$$

$$C_{10} = (R+15) \bmod 27 = (18+15) \bmod 27 = 33 \bmod 27 = 06 = G$$

El criptograma será: $C = \text{HOXJS QSHOG}$.

b) $M_i = (c_i + 27 - 15) \bmod 27 = (c_i + 12) \bmod 27$

$$M_{01} = (H+12) \bmod 27 = (07+12) \bmod 27 = 19 \bmod 27 = 19 = S$$

$$M_{02} = (O+12) \bmod 27 = (15+12) \bmod 27 = 27 \bmod 27 = 00 = A$$

$$M_{03} = (X+12) \bmod 27 = (26+12) \bmod 27 = 38 \bmod 27 = 11 = L$$

$$M_{04} = (J+12) \bmod 27 = (09+12) \bmod 27 = 21 \bmod 27 = 21 = V$$

$$M_{05} = (S+12) \bmod 27 = (19+12) \bmod 27 = 31 \bmod 27 = 04 = E$$

$$M_{06} = (Q+12) \bmod 27 = (17+12) \bmod 27 = 29 \bmod 27 = 02 = C$$

$$M_{07} = (S+12) \bmod 27 = (19+12) \bmod 27 = 31 \bmod 27 = 04 = E$$

$$M_{08} = (H+12) \bmod 27 = (07+12) \bmod 27 = 19 \bmod 27 = 19 = S$$

$$M_{09} = (O+12) \bmod 27 = (15+12) \bmod 27 = 27 \bmod 27 = 00 = A$$

$$M_{10} = (G+12) \bmod 27 = (06+12) \bmod 27 = 18 \bmod 27 = 18 = R$$

El mensaje descifrado es: $M = \text{SALVE CÉSAR}$.

• Cifradores por decimación pura

Ya hemos visto que si la constante de decimación es igual a la unidad, el cifrador se transforma en uno de desplazamiento puro como el del César. Si además la constante de desplazamiento es cero, entonces se transmite en claro lo que no tiene sentido criptográfico. Por el contrario, si la constante de decimación es mayor que la unidad y la constante de desplazamiento es cero, nos encontramos con un cifrador por decimación pura. En este caso las ecuaciones de cifra y descifrado serán:

$$C_i = a * M_i \bmod n \quad \boxed{1.9}$$

$$M_i = a^{-1} * C_i \bmod n \quad \boxed{1.10}$$

donde a^{-1} será el inverso del factor de decimación en el cuerpo n ; es decir $\text{inv}(a, n)$.

Ejemplo 1.16: a) Encuentre el alfabeto de cifrado para la transformación monoalfabética por decimación $C_i = 20 * M_i \bmod 27$.
b) Cifre con este alfabeto el siguiente mensaje $M = \text{DILE A LAURA QUE LA QUIERO}$.

Solución: a)

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

 $C_0 = 20 * A \bmod 27 = 20 * 0 \bmod 27 = 0 = A$
 $C_1 = 20 * B \bmod 27 = 20 * 1 \bmod 27 = 20 = T$
 $C_2 = 20 * C \bmod 27 = 20 * 2 \bmod 27 = 40 \bmod 27 = 13 = N$, etc.
b) El criptograma será $C = \text{GYEZA EAOJA POZEA POYZJ D}$.

Puesto que sólo se aplica este algoritmo de cifra si el valor de la constante de decimación es primo entre sí con el módulo de trabajo, se asegura la existencia del inverso y por tanto la posibilidad de descifrar el criptograma según la ecuación (1.10).

Ejemplo 1.17: Descifre el criptograma que se indica utilizando la ecuación (1.10) si se conoce que éste se ha obtenido con sustitución por decimación pura con una constante de decimación igual a 22.

$C = \text{MÑZHW DHBGR HZZAU DHRG}$.

Solución: El inverso del factor de decimación 22 es $\text{inv}(22, 27) = 16$, luego:

$M_{01} = 16 * 12 \bmod 27 = 192 \bmod 27 = 03 = D$
 $M_{02} = 16 * 14 \bmod 27 = 224 \bmod 27 = 08 = I$
 $M_{03} = 16 * 26 \bmod 27 = 416 \bmod 27 = 11 = L$
 $M_{04} = 16 * 07 \bmod 27 = 112 \bmod 27 = 04 = E$
 $M_{05} = 16 * 23 \bmod 27 = 368 \bmod 27 = 17 = Q$
 $M_{06} = 16 * 03 \bmod 27 = 048 \bmod 27 = 21 = U$
 $M_{07} = 16 * 07 \bmod 27 = 112 \bmod 27 = 04 = E$
 $M_{08} = 16 * 01 \bmod 27 = 016 \bmod 27 = 16 = P$
 $M_{09} = 16 * 06 \bmod 27 = 096 \bmod 27 = 15 = O$
 $M_{10} = 16 * 18 \bmod 27 = 288 \bmod 27 = 18 = R$
 $M_{11} = 16 * 07 \bmod 27 = 112 \bmod 27 = 04 = E$
 $M_{12} = 16 * 26 \bmod 27 = 416 \bmod 27 = 11 = L$
 $M_{13} = 16 * 26 \bmod 27 = 416 \bmod 27 = 11 = L$

$$M_{14} = 16 * 00 \bmod 27 = 000 \bmod 27 = 00 = A$$

$$M_{15} = 16 * 21 \bmod 27 = 336 \bmod 27 = 12 = M$$

$$M_{16} = 16 * 03 \bmod 27 = 048 \bmod 27 = 21 = U$$

$$M_{17} = 16 * 07 \bmod 27 = 112 \bmod 27 = 04 = E$$

$$M_{18} = 16 * 18 \bmod 27 = 288 \bmod 27 = 18 = R$$

$$M_{19} = 16 * 06 \bmod 27 = 096 \bmod 27 = 15 = O$$

El mensaje descifrado es $M = DILE QUE POR ELLA MUERO$

• Cifradores por transformación afín

En los cifradores genéricos, si se cumple que la constante de decimación a es mayor que 1 y la constante de desplazamiento b distinto de cero, hablamos de cifra por transformación afín. Las ecuaciones serán en este caso:

$$C_i = (a * M_i + b) \bmod n \quad 1.11$$

$$M_i = a^{-1} (C_i - b) \bmod n \quad 1.12$$

La ecuación de descifrado también podemos escribirla como sigue:

$$M_i = a^{-1} (C_i + n - b) \bmod n \quad 1.13$$

Ejemplo 1.18: Encuentre el alfabeto de cifrado monoalfabético para la siguiente relación de transformación $C_i = (4 * M_i + 5) \bmod 27$.

Solución:

| | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| M_i | A | B | C | D | E | F | G | H | I | J | K | L | M | N | Ñ | O | P |
| C_i | F | J | N | Q | U | Y | C | G | K | Ñ | R | V | Z | D | H | L | O |

$C_0 = (4 * A + 5) \bmod 27 = (4 * 0 + 5) \bmod 27 = 5 = F$
 $C_1 = (4 * B + 5) \bmod 27 = (4 * 1 + 5) \bmod 27 = 9 = J$
 $C_2 = (4 * C + 5) \bmod 27 = (4 * 2 + 5) \bmod 27 = 13 = N$, etc.

Del ejemplo anterior, observe que el desplazamiento indica dónde comienza la secuencia del alfabeto de cifrado y la decimación los saltos que va dando en el alfabeto original para recorrerlo en su totalidad. ¿Qué sucederá si aplicamos una relación de transformación monoalfabética que no cumpla con las condiciones anteriores? El siguiente ejemplo nos aclarará esta situación..

Ejemplo 1.19: Encuentre el alfabeto de cifrado monoalfabético para la siguiente relación de transformación $C_i = (3 * M_i + 2) \bmod 27$.

Solución: Aplicando la ecuación (1.5) se obtiene:

| | | | | | | | | | | | | | | | | | |
|-------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| M_i | A | B | C | D | E | F | G | H | I | J | K | L | M | N | Ñ | O | P |
| C_i | C | F | I | L | Ñ | Q | T | W | Z | C | F | I | L | Ñ | Q | T | W |

¿Qué sucede en este caso? Con los valores anteriores, no es válida la relación de transformación de alfabetos pues $\text{mcd}(3, 27) \neq 1$ lo que se comprueba en el hecho de que no obtenemos una equivalencia unívoca entre caracteres. Para este caso en que $a = 3$, se va repitiendo el mismo conjunto de 9 ($27/3$) caracteres $CFILÑQTWZ$ por lo que no se obtiene el conjunto completo de restos del módulo y, por tanto, no es

posible utilizarlo como cifrador. De hecho si el criptograma presenta la letra C, no se sabe si corresponde a la acción de cifrar en el texto en claro las letras J, R o A.

Ejemplo 1.20: *Usando un cifrador monoalfabético por decimación y adición según la transformación $(5 \cdot M_i + 8) \bmod 27$, cifre el siguiente mensaje.*

M = DÁBALE ARROZ A LA ZORRA EL ABAD.

Solución:

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

Luego, C = WINIJ BIQQC DIJID CQQIB JINIW.

Al igual que en el cifrador del César, podemos incluir una clave secreta para aumentar la seguridad del sistema. Primero aplicamos la relación de transformación para encontrar un alfabeto de cifrado, luego escribimos la clave a partir de una posición p_0 y finalmente se desplazan los caracteres restantes del alfabeto de cifrado encontrado a partir de la posición final de la clave como se muestra en el siguiente ejemplo.

Ejemplo 1.21: *Aplicando la transformación $C_i = (7 \cdot M_i + 2) \bmod 27$ conjuntamente con la clave K = REFRANERO ESPAÑOL posicionada en $p_0 = 5$, se pide cifrar el mensaje M = NO HAY MAL QUE POR BIEN NO VENGA.*

Solución: *Aplicando la transformación $(7 \cdot M_i + 2) \bmod 27$, obtenemos el siguiente alfabeto de cifrado:*

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

A continuación escribimos la clave en $p_0 = 5$:

_ _ _ _ _ R E F R A N O S P Ñ L _ _ _ _ _

Completando ahora el alfabeto de cifrado desde la posición final de la clave, se obtiene:

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

Luego, C = ÑCFHG PHSWX VJCDU AVÑÑC YVÑEH.

La transformación anterior sigue manteniendo una relación monoalfabética, independientemente de cómo se distribuyan los caracteres; es más, el alfabeto de cifrado no tiene por qué seguir una lógica ni mucho menos una relación matemática como parece deducirse por lo visto; basta con que exista una relación de uno a uno entre alfabeto claro y alfabeto de cifrado. Lo único que logramos con ello es cambiar la relación de correspondencia de caracteres y, en última instancia, ponerle las cosas algo más difíciles a nuestro enemigo criptoanalista, desgraciadamente sólo un poquito.

Puesto que estos cifradores monoalfabéticos genéricos poco se diferencian de los vistos en los apartados anteriores, tendrán la misma fragilidad ante un hipotético ataque de un criptoanalista.

1.4.5. Criptoanálisis de cifrados monoalfabéticos por sustitución

El criptoanálisis de los cifradores monoalfabéticos genéricos por sustitución, esto es aquellos en los que se cumple la *relación afín* $(a \cdot M + b)$, pueden resolverse fácilmente aplicando estadísticas del lenguaje al igual que en el cifrado del César. En

este caso, como además de un desplazamiento b existe una constante de decimación a , podemos plantear un sistema de dos ecuaciones para asignar posibles valores a ambas variables en función del comportamiento estadístico que observamos en los caracteres del criptograma. Como la transformación es $C_i = (a \cdot M_i + b) \bmod n$ asociamos, según la frecuencia relativa de aparición de caracteres en el criptograma, valores de posición de dichos caracteres con los del alfabeto en claro. A continuación haremos un ensayo con diferentes relaciones de congruencia, a partir de unas supuestas correspondencias entre caracteres del criptograma con los caracteres del texto en claro, para ver cómo funciona este método de ataque.

Si sospechamos que la letra más frecuente en un criptograma cualquiera, por ejemplo la letra $M = 12$, se corresponde con la letra más frecuente en el lenguaje castellano, es decir la letra $E = 4$, establecemos la primera relación de equivalencia:

$$a \cdot 4 + b = 12 \bmod 27$$

Siguiendo con la característica de los monogramas en castellano, pensamos ahora que existe una relación directa entre el segundo carácter más frecuente del criptograma, por ejemplo la letra $G = 6$, con la segunda letra más frecuente del lenguaje, $A = 0$. Esto nos da una segunda ecuación:

$$a \cdot 0 + b = 6 \bmod 27$$

Deducimos que $b = 6$. Si reemplazamos este valor en la primera ecuación:

$$\begin{aligned} a \cdot 4 + 6 &= 12 \bmod 27 & \Rightarrow & a \cdot 4 = 6 \bmod 27 \\ a &= 6 \cdot \text{inv}(4, 27) \bmod 27 & \Rightarrow & a = 6 \cdot 7 \bmod 27 \Rightarrow a = 15 \end{aligned}$$

Este resultado no nos sirve pues si $a = 15$ entonces $\text{mcd}(a, n) = 3$ y esto no puede dar lugar a un alfabeto de cifrado pues no genera el CCR(27). A un resultado similar habríamos llegado si, por ejemplo, manteniendo la relación de la primera ecuación hubiésemos supuesto una correspondencia entre la letra $R = 18$ del criptograma y la letra $C = 2$ del texto en claro. Le dejo esto como ejercicio.

Para no aburrirle con esto, vamos a suponer ahora que las correspondencias válidas observadas entre caracteres del criptograma y texto en claro son las siguientes:

| | | |
|--------------------------|--------------------|------------------------------|
| $M = 12$ del criptograma | se corresponde con | $E = 4$ del texto en claro. |
| $T = 20$ del criptograma | se corresponde con | $R = 18$ del texto en claro. |

Establecemos así el siguiente sistema de ecuaciones:

$$\begin{aligned} a \cdot 4 + b &= 12 \bmod 27 \\ a \cdot 18 + b &= 20 \bmod 27 \end{aligned}$$

Restando la primera de la segunda se tiene:

$$a \cdot 14 = 8 \bmod 27 \Rightarrow a = 8 \cdot [\text{inv}(14, 27)] \bmod 27 = 8 \cdot 2 \bmod 27 = 16$$

Sustituyendo $a = 16$ en una de las ecuaciones, obtenemos $b = 2$. Como ahora $\text{mcd}(16, 27) = 1$, entonces el sistema de ecuaciones encontrado podría entregar una solución válida. En este caso la transformación de cifrado aplicada *podría ser*:

$$C_i = (16 \cdot M_i + 2) \bmod 27$$

Si esta congruencia no nos entrega un texto en claro con sentido, tendremos que buscar otras correspondencias entre caracteres del criptograma y del alfabeto, según su distribución de frecuencias, y volver a plantear el sistema de ecuaciones hasta encontrar la transformación que dé lugar al mensaje esperado. Observe que, aunque las relaciones de congruencia sean válidas, no por ello dicha transformación dará lugar a una solución válida en el espacio de los mensajes.

Ejemplo 1.22: *Describe el siguiente criptograma obtenido mediante una transformación monoalfabética por decimación y desplazamiento sin clave: C = NAQÑF EKNDP NCIVU FPUAN EUJIP FCNER NFRÑF UNPLN AFPFQ TFPEI JRTÑE FPKÑI KTAPF LIKIÑ AIPÑU RCUJI PCIVU CUNER IRLNP TJIAF NEOIÑ CFLNC NLUFA TEF.*

Solución: Los caracteres de mayor frecuencia del criptograma anterior son: $F = 14$, $N = 13$ e $I = 12$. Esto nos hace sospechar que se correspondan con las letras A, E y O del alfabeto en claro.

Puesto $n = 27 = 3^3$ entonces $\phi(27) = 3^{3-1}(3-1) = 18$. Haremos sólo dos intentos para mostrar cómo funciona este método de ataque:

1ª Aproximación (que nos lleva a una solución falsa):

Texto claro: $E(4) \Rightarrow$ Criptograma: $F(5)$

Texto claro: $A(0) \Rightarrow$ Criptograma: $N(13)$

Texto claro: $O(15) \Rightarrow$ Criptograma: $I(8)$

Luego: $(a \cdot 4 + b) = 5 \bmod 27$

$(a \cdot 0 + b) = 13 \bmod 27 \Rightarrow b = 13$

$(a \cdot 15 + b) = 8 \bmod 27$

Restando la primera ecuación a la tercera: $a \cdot 11 = 3 \bmod 27$.

Luego, $a = (3 \cdot \text{inv}(11, 27)) \bmod 27 = 3 \cdot 5 \bmod 27 = 15 \Rightarrow a = 15$.

La solución $E(M) = (15M_i + 13)$ se descarta pues $\text{mcd}(15, 27) \neq 1$.

2ª Aproximación (que nos lleva ahora sí a una solución verdadera):

Texto claro: $E(4) \Rightarrow$ Criptograma: $N(13)$

Texto claro: $A(0) \Rightarrow$ Criptograma: $F(5)$

Texto claro: $O(15) \Rightarrow$ Criptograma: $I(8)$

Luego: $(a \cdot 4 + b) = 13 \bmod 27$

$(a \cdot 0 + b) = 5 \bmod 27 \Rightarrow b = 5$

$(a \cdot 15 + b) = 8 \bmod 27$

Restando la primera ecuación a la tercera: $a \cdot 11 = 22 \bmod 27$.

Luego $a = (22 \cdot \text{inv}(11, 27)) \bmod 27 = 22 \cdot 5 \bmod 27 \Rightarrow a = 2$.

Esto da lugar a $E(M) = (2M_i + 5) \bmod 27$, cuyo alfabeto es:

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

Luego el criptograma se describe como:

$M =$ EL GRAN PEZ SE MOVÍA SILENCIOSAMENTE A TRAVÉS DE LAS AGUAS NOCTURNAS, PROPULSADO POR LOS RÍTMICOS

MOVIMIENTOS DE SU COLA EN FORMA DE MEDIA LUNA.
(Primer párrafo del libro "Tiburón", de P. Benchley).

En el caso en que se utilice una clave, el método anterior falla porque deja de existir una relación matemática directa y constante entre el alfabeto en claro y el alfabeto de cifrado. La única solución de ataque, siempre y cuando se conozca o sospeche que se trata de un cifrado monoalfabético, consistirá en buscar digramas, trigramas, y en general formación de palabras características del lenguaje repetidos en el criptograma, evidentemente con caracteres diferentes, y de esta forma obtener el alfabeto de cifrado como ya ha sido comentado en párrafos anteriores. Nos encontramos pues con una tarea más tediosa que la anterior pero, no obstante, relativamente sencilla para este tipo de cifradores.

¿Qué sucede con la distancia de unicidad en estos criptosistemas genéricos? Si analizamos las distintas posibilidades de alfabetos de cifrado, observamos que las posibles transformaciones están ligadas directamente con el factor de decimación. Esto es, la constante de desplazamiento puede tener cualquier valor puesto que asegura una operación inversa; sin embargo, la constante de decimación debe cumplir la condición de ser primo con el módulo. Esto va a indicar que las combinaciones de alfabetos de cifrado serán $n \cdot \phi(n)$; es decir, los n posibles desplazamientos por el indicador de Euler que indica el número de elementos que contiene el CRR(n), es decir primos con el módulo y que por tanto sirven como factor de decimación.

Ejemplo 1.23: *Encuentre la distancia de unicidad de un cifrador genérico de sustitución sin clave para $n = 27$.*

Solución: $N = H(K)/D = \log_2(n \cdot \phi(n))/D = \log_2(27 \cdot 18)/3,4 = 2,6$. Esto es, debemos contar con 3 caracteres cifrados como mínimo.

El valor de la distancia de unicidad para estos cifradores afines es muy bajo puesto que también lo es la cantidad de alfabetos. Con un valor de $n = 27$, solamente existen 486 alfabetos distintos, es decir, 26 desplazamientos por 18 decimaciones. Para hacer crecer este valor podemos usar una clave. En esta situación, el cifrador se convierte en uno monoalfabético con clave, con $27! \approx 10^{28}$ alfabetos diferentes, un valor considerable no cabe duda y que nos entrega una distancia de unicidad igual a 28 caracteres como vimos en el Ejemplo 1.13.

Como hemos comprobado, cualquier cifrador por sustitución monoalfabeto, bien sea por decimación, por desplazamiento o genérico con ambas transformaciones, incluso cuando utilizamos una clave, es muy poco seguro y su ataque en muchos casos se convierte en un paseo para el hipotético criptoanalista. El principal problema de estos cifradores está en que el texto cifrado es un *fiel reflejo* del lenguaje, manifestándose en aquél las características de redundancia del lenguaje. La primera solución que se nos puede ocurrir, que no la óptima por cierto, para evitar que en el criptograma se vea reflejada la redundancia del lenguaje es la utilización de *homófonos*, lo que da lugar a este tipo de cifradores que veremos a continuación.

1.5. CIFRADORES POR HOMÓFONOS

1.5.1. Cifradores por homófonos de primer orden

¿Qué entendemos por *homófonos*? La definición del vocablo puede encontrarse en cualquier diccionario: "*palabras de igual pronunciación o sonido y distinto significado*" como por ejemplo *hola* y *ola*. En criptografía entenderemos por homófonos a las distintas representaciones que damos al mismo carácter sin seguir ninguna relación o función determinada. Por ejemplo, si establecemos una relación entre los 27 caracteres del alfabeto con los 100 primeros números del 00 al 99, podríamos asociar a la letra A los siguientes números: 3, 16, 19, 24, 56, 71, 73, 88, 97. Luego, el receptor autorizado al conocer esta correspondencia simplemente reemplaza dichos números por la letra A para descifrar el mensaje.

Esto da lugar a los denominados *cifradores por homófonos*, cuya característica principal es la de *suavizar* la distribución de frecuencias típica del lenguaje, de forma que el criptoanalista no pueda emplear las técnicas estadísticas vistas en los apartados anteriores. La técnica consiste entonces en asignar un mayor número de homófonos a los caracteres más frecuentes del lenguaje y un menor número a aquellos menos frecuentes, con el objeto de que la distribución de estos valores se asemeje lo más posible a una *distribución normal*.

Observe que en tanto un carácter del texto en claro se cifrará con más de un carácter o símbolo del alfabeto de cifrado, no estamos ya en presencia de un criptosistema monoalfabético. Además, el algoritmo de cifra no tiene porqué seguir una función determinada de asignación de homófonos durante el proceso. Por lo tanto, en este tipo de cifrados hacemos corresponder cada uno de los caracteres del alfabeto del texto en claro con un conjunto de elementos $f(a)$ que denominamos homófonos y que pueden ser cualquier tipo de signos, figuras o números. Luego, si el mensaje M está compuesto por los elementos $\{M_1, M_2, \dots, M_m\}$, entonces se tiene que $\{C_1, C_2, \dots, C_m\}$ será el conjunto de elementos del criptograma en donde cada C_i se toma al azar a partir de un conjunto de homófonos para $f(M_i)$. Esto quiere decir que un mensaje M con una cadena de caracteres $M_1M_2M_3\dots M_m$ se cifrará como se indica:

$$M = M_1M_2M_3\dots M_m \Rightarrow C = f(M_1) f(M_2) f(M_3)\dots f(M_m)$$

Por ejemplo, si se asignan los homófonos que se muestran en la Figura 1.18, podríamos cifrar el mensaje $M = \text{BÁJAME LA JAULA JAIME}$ como se indica:

| | | | | | | | | | |
|--|---|----|----|----|----|----|----|----|----|
| A | ⇒ | 03 | 19 | 36 | 83 | 91 | | | |
| B | ⇒ | 23 | 62 | | | | | | |
| E | ⇒ | 07 | 25 | 28 | 62 | 70 | 88 | 89 | 97 |
| I | ⇒ | 13 | 55 | 70 | | | | | |
| J | ⇒ | 43 | | | | | | | |
| L | ⇒ | 18 | 53 | 60 | | | | | |
| M | ⇒ | 10 | 33 | 71 | 80 | | | | |
| U | ⇒ | 06 | | | | | | | |
| M = B A J A M E L A J A U L A J A I M E | | | | | | | | | |
| C = 62 36 43 03 71 25 18 91 43 19 06 53 83 43 83 55 10 97. | | | | | | | | | |

Figura 1.18. Asignación de homófonos para la cifra que se indica.

No obstante, cualquier criptograma que respete la asignación antes indicada es también válido, pues el receptor autorizado conocerá dicha tabla de homófonos y podrá descifrar el criptograma. Es decir, para el ejemplo anterior, también es válida entre otras la siguiente transformación $C = 2303439133 \ 0718 \ 3643030660 \ 834336131070$.

El cifrador homofónico más importante de la historia de la criptografía es el atribuido al aventurero *Thomas Jefferson Beale*, quien en 1821 deja tres mensajes cifrados, llamados respectivamente B_1 , B_2 y B_3 , en el que supuestamente entrega todas las pistas para descubrir un fabuloso tesoro por él enterrado en Virginia, Estados Unidos. La técnica aplicada por Beale para formar el conjunto de homófonos del cifrado B_2 no deja de ser sorprendente: asigna números a los caracteres del alfabeto según la posición de la palabra en cuestión que comienza con dicha letra dentro del texto de la *Declaración de la Independencia de los Estados Unidos*, cuyas 107 primeras palabras se muestran en la Figura 1.19. Lo siento, tendrá que traducirlo Ud. mismo.

| | |
|--|-----------|
| <i>When, in the course of human events, it becomes necessary</i> | (01-10) |
| <i>for one people to dissolve the political bands which have</i> | (11-20) |
| <i>connected them with another, and to assume among the Powers</i> | (21-30) |
| <i>of the earth the separate and equal station to wich</i> | (31-40) |
| <i>the Laws of Nature and of Nature's God entitle them,</i> | (41-50) |
| <i>a decent respect to the opinions on mankind requires that</i> | (51-60) |
| <i>they should declare the causes wich impel them to the</i> | (61-70) |
| <i>separation. We hold these truths to be self-evident; that</i> | (71-80) |
| <i>all men are created equal, that they are endowed by</i> | (81-90) |
| <i>their Creator with certain unalienable rights; that among these are</i> | (91-100) |
| <i>Life, Liberty, and the pursuit of Happiness.</i> | (101-107) |

Figura 1.19. Comienzo de la Declaración de Independencia de los Estados Unidos.

Por ejemplo, siguiendo el texto de la Declaración de la Independencia de los Estados Unidos mostrado en la figura anterior, se obtienen los homófonos de valor menor que 100 para las letras A, B, C, D y E que se recogen en la Figura 1.20.

| | | | | | | | | | | | |
|-----------------|----|----|----|----|----|----|----|----|----|----|----|
| A \Rightarrow | 24 | 25 | 27 | 28 | 36 | 45 | 51 | 81 | 83 | 88 | 98 |
| B \Rightarrow | 9 | 18 | 77 | 90 | | | | | | | |
| C \Rightarrow | 4 | 21 | 65 | 84 | 92 | 94 | | | | | |
| D \Rightarrow | 15 | 52 | 63 | | | | | | | | |
| E \Rightarrow | 7 | 33 | 37 | 49 | 79 | 85 | 89 | | | | |

Figura 1.20. Algunos homófonos del cifrado B_2 de Beale.

Así, el cifrado B_2 que especificaba el valor del tesoro y cómo había sido enterrado, comenzaba de la siguiente forma: "*I have deposited...*" y terminaba con el siguiente mensaje: "*Paper Number One describes the exact locality of the vault, so that no difficulty will be hand in finding it*" (su traducción es elemental). Esto trajo de cabeza a criptoanalistas aficionados cuya mayor ilusión era encontrar dicho tesoro. Tras diversos estudios más serios por parte del *Laboratorio de Criptografía George Fabyan* en Riverbank Illinois, se llega a la conclusión de que B_1 sigue el mismo principio de cifrado que B_2 pero, por muchos intentos que se hacen, no se llega a describirlo.

Ultimamente, *James J. Gillogly* y *Louis Kruh*, exponen en 1980 y 1982, respectivamente, en sendos artículos de la revista *Cryptologia*³, las anomalías encontradas en el primer criptograma de Beale, llegando a la conclusión de que se trata de una gran broma, posiblemente llevada a cabo por *James B. Ward*, vecino de Campbell County en Virginia, a quien supuestamente habían llegado los criptogramas de mano de *Robert Morris*, el tabernero a quien Beale habría confiado su secreto al abandonar el pueblo... algo enrevesado, lo reconozco, pero que no deja de ser curioso.

El problema de la generación de homófonos a partir de un texto está en que, salvo que éste tenga una extensión muy grande, no se consiguen homófonos para algunas letras pocas frecuentes como inicios de palabras, como sería el caso de las letras K, Ñ y W para el castellano. La única solución consistiría en dejar unos números al final del cuerpo de homófonos para estos caracteres poco frecuentes.

Ejemplo 1.24: *Construya una tabla de homófonos con las 99 primeras palabras del libro "Cien años de soledad" de Gabriel García Márquez y luego cifre el siguiente mensaje: M = UNA GRAN NOVELA.*

Solución: *El texto indicado es el siguiente:*

| | |
|---|----|
| "Muchos años después, frente al pelotón de fusilamiento, el coronel | 10 |
| Aureliano Buendía había de recordar aquella tarde remota en que | 20 |
| su padre lo llevó a conocer el hielo. Macondo era | 30 |
| entonces una aldea de veinte casas de barro cañabrava construidas | 40 |
| a la orilla de un río de aguas diáfanas que | 50 |
| se precipitaban por un lecho de piedras pulidas, blancas y | 60 |
| enormes como huevos prehistóricos. El mundo era tan reciente, que | 70 |
| muchas cosas carecían de nombre, y para mencionarlas había que | 80 |
| señalarlas con el dedo. Todos los años, por el mes | 90 |
| de marzo, una familia de gitanos desarrapados plantaba su ..." | 99 |

Los caracteres con homófonos serán en este caso:

| | | | | | | | | | | | | | | |
|---|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| A | 2 | 5 | 11 | 16 | 25 | 33 | 41 | 48 | 87 | | | | | |
| B | 12 | 38 | 59 | | | | | | | | | | | |
| C | 10 | 26 | 36 | 39 | 40 | 62 | 72 | 73 | 82 | | | | | |
| D | 3 | 7 | 14 | 34 | 37 | 44 | 47 | 49 | 56 | 74 | 84 | 91 | 95 | 97 |
| E | 9 | 19 | 27 | 30 | 31 | 61 | 65 | 67 | 83 | 89 | | | | |
| F | 4 | 8 | 94 | | | | | | | | | | | |
| G | 96 | | | | | | | | | | | | | |
| H | 13 | 28 | 63 | 79 | | | | | | | | | | |
| L | 23 | 24 | 42 | 55 | 86 | | | | | | | | | |
| M | 1 | 29 | 66 | 71 | 78 | 90 | 92 | | | | | | | |
| N | 75 | | | | | | | | | | | | | |
| O | 43 | | | | | | | | | | | | | |
| P | 6 | 22 | 52 | 53 | 57 | 58 | 64 | 77 | 88 | 98 | | | | |
| Q | 20 | 50 | 70 | 80 | | | | | | | | | | |
| R | 15 | 18 | 46 | 69 | | | | | | | | | | |
| S | 21 | 51 | 81 | 99 | | | | | | | | | | |

³ Deavours Cipher, Kahn David, Kruh Louis, Mellen Greg, Winkel Brian, "*Cryptology: Machines, History & Methods*", Artech House, 1989, pp. 491-505.

| | | | | |
|---|----|----|----|----|
| T | 17 | 68 | 85 | |
| U | 32 | 45 | 54 | 93 |
| V | 35 | | | |
| Y | 60 | 76 | | |

Uno de los tantos criptogramas podría ser:

$C = 327511\ 96463375\ 754335305548$.

Como puede apreciar, en el ejemplo anterior no se ha podido encontrar homófonos para todo el alfabeto por lo que no podríamos cifrar, por ejemplo, el mensaje $M = \text{Una obra magistral}$.

Además de lo anterior, si bien el método utilizado por Beale para la generación de homófonos entrega un cifrado que es difícil romper, cumple sólo parcialmente con el principio básico de estos cifradores, cual es la destrucción de la distribución característica de los caracteres a través de una distribución plana de los mismos en el criptograma. Esto es, si en un determinado lenguaje (castellano por ejemplo en módulo 27) las letras P, U, R y A presentan unos valores aproximados de frecuencia relativa iguales a 3, 4, 7 y 11 por ciento, respectivamente, entonces sobre 100 números o signos elegidos como homófonos, asignaríamos 3 homófonos a la letra P, 4 a la letra U, 7 a la R y 11 a la letra A. En el método propuesto por Beale, no se consigue esta distribución de homófonos.

De los ejemplos anteriores, ninguno de los dos textos tomados como referencia para los homófonos -la Declaración de la Independencia de los Estados Unidos en el primero y el libro de García Márquez en el segundo- cumplen con esto. Entre otras diferencias notables, en ambos casos la letra A está por encima de la letra E, lo que no se corresponde ni con el lenguaje inglés ni el castellano. La única ventaja que tiene elegir estos textos como generadores de homófonos está en la sencillez del algoritmo de asignación y he aquí, precisamente, su gran debilidad puesto que el criptoanalista puede llegar a descubrir *toda la clave* por intuición a partir de un pequeño trozo por todos conocidos. Por ejemplo, está claro que el texto "*En un lugar de La Mancha ...*" es una malísima elección puesto que es una pista excelente para un probable criptoanalista. Un sistema de homófonos con una mayor seguridad sería, por ejemplo, asignar números de tres dígitos a las letras del alfabeto, obteniendo dichos números a partir de una página en particular de una Guía de Teléfonos; si es de otro país y antigua mejor aún. Dicha posición o página sería la clave secreta del criptosistema en cuestión.

1.5.2. Cifradores por homófonos de orden mayor

A mayor cantidad de texto cifrado existe una mayor facilidad para abordar el criptoanálisis. La razón es que una única clave descifra el criptograma C en un texto con sentido en el espacio de los mensajes M . Esto se ve agravado si la clave está asociada con algún documento o texto ampliamente conocido como ya se ha comentado.

El método de cifrado con homófonos de mayor orden intenta solucionar este problema. La idea es que un mismo criptograma pueda ser descifrado o, mejor aún, descriptado con claves diferentes y produzcan en cada caso un mensaje con sentido en

el espacio M con igual probabilidad. Recuerde que este tipo de soluciones eran consideradas como falsas según el modelo de cifrador aleatorio de Hellman visto en un capítulo anterior.

• Cifrador homofónico de segundo orden

En este cifrador se realiza la asignación de homófonos mediante una cuadrícula de forma que dicho valor representa a una letra si se lee a través de las filas y otra letra distinta si la lectura se hace a través de las columnas. Así, enviamos el mensaje verdadero y uno falso, ambos cifrados con el valor de dicha cuadrícula, de forma que el criptoanalista en el mejor de los casos podrá contar con dos mensajes, sin saber cuál de ellos es el verdadero. El algoritmo es el siguiente:

- Los números 1 al n^2 se distribuyen de forma aleatoria en una matriz K de orden $n \times n$, cuyas filas y columnas corresponden a los caracteres del alfabeto.
- Para cada carácter a del alfabeto, la fila de la matriz K define un conjunto de homófonos $f_F(a)$ y la columna define otro conjunto de homófonos $f_C(a)$.
- Para proceder al cifrado, se escriben los dos mensajes, uno debajo del otro, el verdadero que llamaremos M y uno falso que denominaremos X . El homófono que se envía como elemento del criptograma es el valor que aparece en la matriz K , en la intersección entre la fila del carácter en claro verdadero y la columna del carácter del mensaje falso.
- Con esto, el criptograma se forma mediante el valor de las siguientes intersecciones $f_F(M_1)f_C(X_1)$, $f_F(M_2)f_C(X_2)$, ..., $f_F(M_m)f_C(X_m)$, en donde f_F es la función lectura en filas y f_C la lectura en columnas.

En la Figura 1.21 se muestra parte de una hipotética tabla de asignación de homófonos para un cifrador con estas características.

| | A | B | C | D | E | F | G | H | I | J | ... |
|-----|----|----|----|----|-----|----|----|----|----|----|-----|
| A | 60 | 47 | 13 | 37 | 5 | 91 | 33 | 19 | 92 | 80 | |
| B | 39 | 8 | 53 | 72 | 9 | 89 | 57 | 93 | 38 | 54 | |
| C | 73 | 1 | 21 | 94 | 65 | 10 | 82 | 58 | 36 | 18 | |
| D | 12 | 48 | 2 | 84 | 20 | 59 | 3 | 11 | 55 | 99 | |
| E | 40 | 88 | 97 | 26 | 52 | 71 | 79 | 35 | 64 | 56 | |
| F | 14 | 95 | 66 | 22 | 83 | 78 | 16 | 41 | 34 | 27 | |
| G | 96 | 85 | 15 | 69 | 25 | 51 | 42 | 76 | 17 | 23 | |
| H | 4 | 61 | 28 | 46 | 100 | 24 | 98 | 70 | 67 | 90 | |
| I | 74 | 29 | 77 | 86 | 50 | 62 | 6 | 43 | 44 | 32 | |
| J | 7 | 49 | 68 | 30 | 45 | 75 | 63 | 87 | 31 | 81 | |
| ... | | | | | | | | | | | |

Figura 1.21. Posible asignación de homófonos para un cifrador de segundo orden.

Luego, al cifrar el mensaje $M = CEDIDA$ con el mensaje falso $X = FIJADA$ con la tabla anterior se obtiene $C = 10\ 64\ 99\ 74\ 84\ 60$. Observe que en estos cifrados el tamaño del texto en claro debe ser el mismo que el del texto falso.

1.5.3. Criptoanálisis de los cifrados por homófonos

Los criptosistemas con homófonos pueden llegar a ser extremadamente difíciles de romper, especialmente si la asignación de tales valores no sigue una lógica como en los ejemplos anteriores, en que éstos eran obtenidos a partir de un texto muy conocido. Con una gran cantidad de texto cifrado es posible encontrar algunas cadenas de números o símbolos que se repiten y, por tanto, se pueden formar digramas, trigramas y en el mejor de los casos palabras y frases completas. Si todo va bien, con un *poquitín* de suerte podremos obtener en algunos casos la tabla de homófonos.

Para los cifradores por homófonos de segundo orden, una técnica que puede dar algún fruto, también en función de que se cuente con una gran cantidad de texto cifrado, consiste en asociar los números repetidos a pares de letras de alta frecuencia, ir rellenando la matriz y, a su vez, buscar digramas, trigramas, palabras, etc., con el objeto de obtener la matriz de cifrado. Análogamente, lo mismo puede decirse para cifradores de mayor orden. No profundizaremos en este tipo de criptoanálisis en este libro pues es menester contar con un texto cifrado muy grande y no tiene sentido ocupar páginas en ello. El lector interesado en las técnicas para romper estos cifradores puede consultar las referencias anteriores como Deavours, Khan y otros.

Ejemplo 1.25: *Haciendo uso de nuestras habilidades y fuentes de información que no vamos a desvelar en este momento, hemos encontrado el siguiente trozo de una tabla de homófonos, relacionada con el criptograma de 43 elementos que se indica. Encuentre los mensajes que han dado lugar al criptograma que se indica.*

C = 699 289 492 124 005 693 404 017 126 559 710 590 700
 258 046 124 200 705 159 200 545 581 545 644 503 388
 590 219 150 041 480 727 086 346 468 603 444 013 668
 077 590 100 301.

Parte de la Tabla de Homófonos (ordenada numéricamente)

1^{er} carácter (Fila) 2^o carácter (Columna).

| | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| 005 | 013 | 017 | 041 | 046 | 077 | 086 | 100 | 124 |
| EL | RE | EA | ED | DL | TC | AV | ES | AA |
| 126 | 150 | 159 | 200 | 219 | 258 | 289 | 301 | 346 |
| AP | TO | UR | SE | AT | OE | EO | SA | NU |
| 388 | 404 | 444 | 468 | 480 | 492 | 503 | 545 | 559 |
| TA | VL | EV | UE | GO | LV | SI | AN | PE |
| 581 | 590 | 603 | 644 | 668 | 693 | 699 | 700 | 705 |
| LU | RA | ML | EC | OA | ZE | PN | TP | YZ |
| 710 | 727 | | | | | | | |
| ON | IY | | | | | | | |

Solución: *leyendo los primeros caracteres de los homófonos del criptograma, es decir filas, tenemos:*

699=PN \Rightarrow P 289=EO \Rightarrow E 492=LV \Rightarrow L 124=AA \Rightarrow A 005=EL \Rightarrow E
 693=ZE \Rightarrow Z 404=VL \Rightarrow V 017=EA \Rightarrow E 126=AP \Rightarrow A 559=PE \Rightarrow P
 710=ON \Rightarrow O 590=RA \Rightarrow R 700=TP \Rightarrow T 258=OE \Rightarrow O 046=DL \Rightarrow D
 124=AA \Rightarrow A 200=SE \Rightarrow S 705=YZ \Rightarrow Y 159=UR \Rightarrow U 200=SE \Rightarrow S

545=AN \Rightarrow A 581=LU \Rightarrow L 545=AN \Rightarrow A 644=EC \Rightarrow E 503=SI \Rightarrow S
 388=TA \Rightarrow T 590=RA \Rightarrow R 219=AT \Rightarrow A 150=TO \Rightarrow T 041=ED \Rightarrow E
 480=GO \Rightarrow G 727=IY \Rightarrow I 086=AV \Rightarrow A 346=NU \Rightarrow N 468=UE \Rightarrow U
 603=ML \Rightarrow M 444=EV \Rightarrow E 013=RE \Rightarrow R 668=OA \Rightarrow O 077=TC \Rightarrow T
 590=RA \Rightarrow R 100=ES \Rightarrow E 301=SA \Rightarrow S

De esta forma se obtiene el mensaje por filas:

M_{Fila} = PELÁEZ, VE A POR TODAS Y USA LA ESTRATEGIA NÚMERO TRES.

Leyendo los segundos caracteres de los homófonos del criptograma, tenemos:

699=PN \Rightarrow N 289=EO \Rightarrow O 492=LV \Rightarrow V 124=AA \Rightarrow A 005=EL \Rightarrow L
 693=ZE \Rightarrow E 404=VL \Rightarrow L 017=EA \Rightarrow A 126=AP \Rightarrow P 559=PE \Rightarrow E
 710=ON \Rightarrow N 590=RA \Rightarrow A 700=TP \Rightarrow P 258=OE \Rightarrow E 046=DL \Rightarrow L
 124=AA \Rightarrow A 200=SE \Rightarrow E 705=YZ \Rightarrow Z 159=UR \Rightarrow R 200=SE \Rightarrow E
 545=AN \Rightarrow N 581=LU \Rightarrow U 545=AN \Rightarrow N 644=EC \Rightarrow C 503=SI \Rightarrow I
 388=TA \Rightarrow A 590=RA \Rightarrow A 219=AT \Rightarrow T 150=TO \Rightarrow O 041=ED \Rightarrow D
 480=GO \Rightarrow O 727=IY \Rightarrow Y 086=AV \Rightarrow V 346=NU \Rightarrow U 468=UE \Rightarrow E
 603=ML \Rightarrow L 444=EV \Rightarrow V 013=RE \Rightarrow E 668=OA \Rightarrow A 077=TC \Rightarrow C
 590=RA \Rightarrow A 100=ES \Rightarrow S 301=SA \Rightarrow A

Ahora se obtiene el mensaje por columna:

$M_{columna}$ = NO VALE LA PENA PELÁEZ, RENUNCIA A TODO Y VUELVE A CASA.

En el ejemplo anterior, incluso conociendo de qué va el *affaire* de nuestro querido amigo *Peláez*, no seremos capaces de dilucidar cuál de los dos mensajes que hemos criptoanalizado será verdadero y cuál falso, salvo que conozcamos la clave de lectura en la tabla. Por lo tanto, nuestro esfuerzo en romper la tabla de homófonos nos ha llevado a un callejón sin salida: tener mucho texto con sentido pero asociado a una gran incertidumbre sobre la información que contiene y la veracidad del mismo. En otras palabras y aunque parezca un sarcasmo, después de *rompernos la cabeza*, *no tenemos nada*.

1.6. CIFRADORES POR SUSTITUCIÓN MONOGRÁMICA POLIALFABETO

Los criptosistemas monoalfabéticos por sustitución y los de transposición o permutación presentan una gran debilidad al poder romperse en muchos casos los criptogramas aplicando unas técnicas sencillas de estadísticas del lenguaje. Esta debilidad está asociada directamente al hecho de que en el primero de ellos la sustitución se realizaba siempre con un único carácter del alfabeto de cifrado y, en el segundo, a que las letras del criptograma serán exactamente las mismas que las del texto en claro y sólo se han roto algunas adyacencias de caracteres. Ambos escenarios entregan una ayuda inapreciable al hipotético criptoanalista. Para solucionar estos problemas, aparecen los cifradores por *sustitución polialfabéticos* que, como su nombre lo indica, usan más de un alfabeto para cifrar. En el fondo, hacen uso de la característica ya apuntada por Alberti y que hemos comentado al comienzo de este capítulo en el apartado 1.1.1.

Los algoritmos de sustitución polialfabética tienen por objeto producir una distribución plana de la frecuencia relativa de los caracteres en el criptograma, de una manera similar a la técnica de los homófonos. Para ello utilizan sustituciones múltiples de forma que en un texto largo se combinan las altas frecuencias de algunos caracteres con otros de menor frecuencia. En otras palabras, si por ejemplo la letra *A*, de alta frecuencia en el lenguaje castellano, se cifra algunas veces como la letra *O* y otras veces como la letra *J* (una de alta frecuencia y la otra de baja) y lo mismo ocurre para la letra *W*, de baja frecuencia en el lenguaje, el efecto final es suavizar la mencionada distribución de frecuencia de todos los caracteres del criptograma.

La técnica anterior consiste en aplicar dos o más alfabetos de cifrado de forma que cada uno de ellos sirva para cifrar los caracteres del texto en claro, dependiendo de la posición relativa de éstos en dicho texto. Por ejemplo, si utilizamos un alfabeto A_1 para cifrar los caracteres de las posiciones impares del mensaje y otro alfabeto A_2 para los caracteres en posiciones pares, entonces en un texto lo suficientemente extenso se tendrá que, aproximadamente, sólo en la mitad de las ocasiones las letras repetidas del texto en claro se repiten como un mismo elemento c_i en el criptograma, lográndose por tanto el efecto de enmascaramiento de la distribución de frecuencia de los monogramas característica del lenguaje.

1.6.1. Cifradores polialfabéticos periódicos

Los sistemas por sustitución polialfabética tienen, por lo general, un período que vendrá dado por la longitud de la clave de cifrado. La única excepción la encontramos en los denominados cifradores polialfabéticos de clave continua, siendo un ejemplo característico el cifrador de Vernam a estudiar en el apartado 1.6.8 y que, como su nombre lo indica, posee una clave tanto o más larga que el texto en claro por lo que serán cifradores no periódicos.

Luego, si dependiendo de la longitud de la clave tenemos d alfabetos de cifrado, habrá una periodicidad en los elementos del criptograma indicada por la siguiente cadena:

$$C = C_1 \dots C_d C_{d+1} \dots C_{2d} C_{2d+1} \dots C_{3d} \dots \quad 1.14$$

O lo que es lo mismo, si $f: A \rightarrow C$ es una aplicación de correspondencia del alfabeto A del texto en claro con el alfabeto de cifrados C_i , con $1 \leq i \leq d$, entonces tendremos que el mensaje $M = M_1 \dots M_d M_{d+1} \dots M_{2d} \dots$, se cifrará repitiendo la secuencia $f_1 \dots f_d$ cada d caracteres. Por lo tanto, podemos concluir que:

$$E_K(M) = f_1(M_1) \dots f_d(M_d) f_1(M_{d+1}) \dots f_d(M_{2d}) f_1(M_{2d+1}) \dots \quad 1.15$$

1.6.2. Cifrador de Vigenère

El cifrador polialfabético más conocido es el sistema de Vigenère, así denominado en honor al criptólogo francés *Blaise de Vigenère* (1523-1596). El sistema utiliza el mismo método que el cifrador del César, esto es una sustitución monográfica

por desplazamiento de k caracteres en el texto, con la diferencia de que dicho desplazamiento viene indicado por el valor numérico asociado a uno de los caracteres de una clave que se *escribe cíclicamente bajo* el mensaje como se indica a continuación:

TEXTO: E N U N L U G A R D E L A M A N C H A D E C U Y O N O M B R E ...
CLAVE: C E R V A N T E S C E R V A N T E S C E R V A N T E S C E R V ...

Según lo anterior, la clave utilizada será *CERVANTES* y tendrá una periodicidad igual a 9, pues son los caracteres que forman esta palabra. Luego, al primer carácter del texto en claro (*E*) se le aplica un desplazamiento equivalente al primer carácter de la clave (*C*), dando lugar a $E + C = (4 + 2) \bmod 27 = 6 = G$; el segundo carácter (*N*) se cifra sumándolo con el segundo carácter de la clave (*E*), $N + E = (13 + 4) \bmod 27 = 17 = Q$, etc. El resultado final será el criptograma: *C = GQMIL HZEKF ICMVN GGZCH VXULI*. Tómese ahora un merecido descanso y compruebe este resultado.

Ejemplo 1.26: *Aplicando relaciones de congruencia como las indicadas en el párrafo anterior, cifre el siguiente mensaje según el método de Vigenère.*

M = DESASTRE NUCLEAR EN MURUROA. Clave K = SOS.

Solución:

| | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| D | E | S | A | S | T | R | E | N | U | C | L | E | A | R | E | N | M | U | R | U | R | O | A |
| S | O | S | S | O | S | S | O | S | S | O | S | S | O | S | S | O | S | S | O | S | S | O | S |
| D+S | E+O | S+S | A+S | S+O | T+S | R+S | E+O | N+S | U+S | C+O | L+S | E+O | A+S | R+O | E+O | N+S | M+S | U+S | R+S | U+S | R+S | O+S | A+S |
| = (3+19) | = (4+15) | = (19+19) | = (0+19) | = (19+15) | = (20+19) | = (18+19) | = (4+15) | = (13+19) | = (21+19) | = (2+15) | = (11+19) | = (4+15) | = (0+19) | = (15+15) | = (13+19) | = (12+19) | = (21+19) | = (18+19) | = (4+19) | = (21+19) | = (18+19) | = (15+15) | = (0+19) |
| mod27 = 22 | mod27 = 19 | mod27 = 11 | mod27 = 19 | mod27 = 7 | mod27 = 12 | mod27 = 10 | mod27 = 19 | mod27 = 5 | mod27 = 13 | mod27 = 17 | mod27 = 3 | mod27 = 19 | mod27 = 19 | mod27 = 6 | mod27 = 13 | mod27 = 4 | mod27 = 13 | mod27 = 10 | mod27 = 23 | mod27 = 13 | mod27 = 10 | mod27 = 3 | mod27 = 19 |
| $\Rightarrow V$ | $\Rightarrow S$ | $\Rightarrow L$ | $\Rightarrow S$ | $\Rightarrow H$ | $\Rightarrow M$ | $\Rightarrow K$ | $\Rightarrow S$ | $\Rightarrow F$ | $\Rightarrow N$ | $\Rightarrow Q$ | $\Rightarrow D$ | $\Rightarrow S$ | $\Rightarrow S$ | $\Rightarrow G$ | $\Rightarrow N$ | $\Rightarrow E$ | $\Rightarrow N$ | $\Rightarrow K$ | $\Rightarrow W$ | $\Rightarrow N$ | $\Rightarrow K$ | $\Rightarrow D$ | $\Rightarrow S$ |

Luego, se obtiene el siguiente criptograma:

C = VSLSH MKSFN QDWOK WBENG NKDS.

Observe que letras repetidas del texto en claro se cifran de forma distinta, dependiendo de su posición relativa respecto a la clave. Es el caso de la letra *E* que se cifra dos veces como *S* al coincidir con la letra *O* de la clave y dos veces como *W* cuando la letra de la clave es *S*. Algo similar ocurre con las letras *A*, *N* y *R* y no así con la *U* que se cifra tres veces como *N*. Por otra parte, una letra repetida del criptograma puede provenir de caracteres distintos del texto en claro. Es el caso de la letra *D* que proviene de los caracteres *L* y *O* del mensaje. Las observaciones anteriores pueden generalizarse teniendo en cuenta el número de alfabetos utilizados en función de la clave. En nuestro ejemplo, si bien la clave *SOS* implica una periodicidad igual a tres, solamente utilizamos dos alfabetos, el correspondiente a la letra *S* y el de la letra *O*.

CLAVE

TEXTO EN CLARO

| | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 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 |
| 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 | 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 | 1 | 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 | 2 | 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 | 3 | 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 | 4 | 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 | 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 |
| G | 6 | 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 | 7 | 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 | 8 | 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 | 9 | 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 | 10 | 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 | 11 | 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 | 12 | 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 | 13 | 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 | 14 | 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 | 15 | 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 | 16 | 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 | 17 | 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 | 18 | 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 | 19 | 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 | 20 | 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 | 21 | 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 | 22 | 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 | 23 | 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 | 24 | 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 | 25 | 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 | 26 | 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 |

Figura 1.22. Tabla de cifrar de Vigenère.

El algoritmo de Vigenère utiliza permutaciones para cifrar los caracteres del texto en claro con una clave. Si extendemos el número de permutaciones hasta su límite superior, en nuestro caso 27, se obtiene la denominada *Tabla de Vigenère* que se muestra en la Figura 1.22.

Para cifrar un texto utilizando la Tabla de Vigenère se procede de la siguiente manera: nos posicionamos en el carácter del texto en claro a cifrar en la primera fila de la tabla y buscamos la letra de la clave en cuestión en la primera columna de la tabla. El elemento C_i del criptograma será la letra de la retícula de intersección entre fila y columna. Por ejemplo la letra *M* cifrada con la clave *O* nos dará el criptograma *A*. Compruebe que al mismo resultado se llega si se cifra el texto en claro *O* con la clave *M*. La primera fila, la de la clave *A*, corresponde a la del texto en claro pues es lo que se obtiene al aplicar un desplazamiento igual a cero.

Ejemplo 1.27: *Utilizando la Tabla de Vigenère y la clave $K = WINDOWS$, cifre el siguiente mensaje: $M = MARIPURI, APAGA ESE ORDENADOR.$*

Solución:
 M A R I P U R I A P A G A E S E O R D E N A D O R
 W I N D O W S W I N D O W S W I N D O W S W I N D

Buscando en la tabla, obtenemos el criptograma:
 $C = IIELE QKEIC DUWWO MBURA FWLBU.$

En el ejemplo anterior la clave tenía longitud 7 aunque sólo hemos hecho uso de 6 alfabetos de cifrado, los de las letras no repetidas de la clave *W, I, N, D, O, S*.

Puesto que la clave está formada por un conjunto de d caracteres, $K = k_1 \dots k_d$, en donde k_i ($i = 1, \dots, d$) entrega la cantidad de desplazamiento del alfabeto i ésimo, la función de transformación de Vigenère para cifrar vendrá dada por:

$$C_i = E_{k_i}(M_i) = (M_i + k_i) \bmod n \quad 1.16$$

Para realizar la operación de descifrado con la tabla se procede de manera inversa. En la fila del carácter i ésimo de la clave, nos posicionamos en la retícula de la letra del criptograma; hecho esto, subimos por esta columna hasta la fila primera del texto en claro y leemos el carácter. Por ejemplo, si el elemento C_i es la letra G y el elemento k_i es la letra \tilde{N} , el resultado será el texto en claro S .

Ejemplo 1.28: *En la sala de mandos se recibe el siguiente criptograma que se sabe ha sido cifrado mediante Vigenère con la clave TORA. Se pide descifrarlo.*

$C = \text{RODAF DLOIG UEGOR TTQRR JSRRE VRRUD J.}$

Solución: *Aplicando el método comentado, obtenemos el inquietante mensaje:*
 $M = \text{YAMAMOTO ORDENA ATACAR PEARL HARBOR.}$

De acuerdo con la operación de cifra indicada en la ecuación (1.16), la función de descifrado de Vigenère deberá utilizar el inverso del desplazamiento aplicado, dando lugar a la expresión:

$$M_i = D_{k_i}(C_i) = (C_i - k_i) \bmod n \quad 1.17$$

Por ejemplo, para el primer elemento del criptograma del ejemplo anterior (R), con letra de clave (T) obtenemos: $(R-T) \bmod 27 = (18-20) \bmod 27 = -2 \bmod 27 = 25 = Y$. Continúe Ud. con el resto del criptograma.

1.6.3. Cifrador autoclave

Es una variante del algoritmo de Vigenère, conocido también como Segundo Cifrado de Vigenère, cuya característica radica en que se cifra el mensaje con una clave que consiste en el mismo mensaje al que se le añade al comienzo una clave denominada *primaria*. Luego, la secuencia de clave será tan larga como el propio mensaje. Por ejemplo, si utilizando la clave $K = \text{MARKETING}$ deseamos cifrar el mensaje $M = \text{YA ES PRIMAVERA EN EL CORTE INGLÉS}$, usando la Tabla de Vigenère obtenemos el siguiente criptograma:

```

M = Y A E S P R I M A V E R A E N E L C O R T E I N G L E S
K = M A R K E T I N G Y A E S P R I M A V E R A E N E L C O
C = K A V C T L P Y G T E V S T E M W C K V L E M Z K V G H

```

La operación de descifrado, conociendo la clave *MARKETING* es igual que en Vigenère. Esto es, siguiendo la ecuación (1.17), desciframos los nueve primeros caracteres como se indica:

$$\begin{aligned}
 K - M &\Rightarrow (10 - 12) \bmod 27 = 25 &\Rightarrow Y \\
 A - A &\Rightarrow (0 - 0) \bmod 27 = 0 &\Rightarrow A \\
 V - R &\Rightarrow (22 - 18) \bmod 27 = 4 &\Rightarrow E
 \end{aligned}$$

C - K $\Rightarrow (2 - 10) \bmod 27 = 19 \Rightarrow S$
 T - E $\Rightarrow (20 - 4) \bmod 27 = 16 \Rightarrow P$
 L - T $\Rightarrow (11 - 20) \bmod 27 = 18 \Rightarrow R$
 P - I $\Rightarrow (16 - 8) \bmod 27 = 8 \Rightarrow I$
 Y - N $\Rightarrow (25 - 13) \bmod 27 = 12 \Rightarrow M$
 G - G $\Rightarrow (6 - 6) \bmod 27 = 0 \Rightarrow A$

Con este método seremos capaces de descifrar sólo los primeros 9 caracteres del criptograma *KAVCTLPYG* correspondientes a la clave primaria *MARKETING*. Para continuar descifrando, hacemos uso de los 9 caracteres ya descifrados (*YAESPRIMA*) que, según el método, irán a continuación de dicha clave. Por lo tanto, desciframos ahora el criptograma *TEVSTEMWC* con la clave *YAESPRIMA* y así sucesivamente hasta obtener el mensaje original ... que se lo dejo como ejercicio. Aunque impresione más que otras técnicas de cifra, el secreto de este criptosistema reside únicamente en el de la clave. Conocida ésta, el criptoanálisis es elemental.

Ejemplo 1.29: Si la clave usada en un cifrador autoclave es *PIZZA*, descifre el siguiente criptograma.

C = SWMCE HHGDI OLXCV OMSGC WXQVO MSGKX TSQDT MEFL

Solución:

| | | |
|-----------------|--------------|----------------------------------|
| Bloque 1: SWMCE | Clave: PIZZA | $\Rightarrow M_1 = \text{DONDE}$ |
| Bloque 2: HHGDI | Clave: DONDE | $\Rightarrow M_2 = \text{ESTAE}$ |
| Bloque 3: OLXCV | Clave: ESTAE | $\Rightarrow M_3 = \text{LSECR}$ |
| Bloque 4: OMSGC | Clave: LSECR | $\Rightarrow M_4 = \text{ETOEL}$ |
| Bloque 5: WXQVO | Clave: ETOEL | $\Rightarrow M_5 = \text{SECRE}$ |
| Bloque 6: MSGKX | Clave: SECRE | $\Rightarrow M_6 = \text{TOEST}$ |
| Bloque 7: TSQDT | Clave: TOEST | $\Rightarrow M_7 = \text{AENLA}$ |
| Bloque 8: MEFL | Clave: AENL | $\Rightarrow M_8 = \text{MASA}$ |

Luego, incluyendo los signos de puntuación, este sabroso mensaje es:

M = ¿DÓNDE ESTÁ EL SECRETO? EL SECRETO ESTÁ EN LA MASA.

(He incluido los signos)

1.6.4. Cifrador de Beaufort

En 1710, *Giovanni Sestri*, basado en el método de cifra de Vigenère, propone un algoritmo simétrico que sirve tanto para cifrar como para descifrar. El invento del cifrador en cuestión finalmente se le atribuye al inglés Sir *Francis Beaufort*, amigo de Sestri, y recibe precisamente el nombre de cifrador de Beaufort. Eso sí que es un amigo. La sustitución empleada en este cifrador sigue la siguiente expresión:

$$C_i = E_{k_i}(M_i) = (k_i - M_i) \bmod n \quad 1.18$$

La Figura 1.23 recoge todos los valores de la Tabla de Beaufort para el lenguaje castellano.

CLAVE

TEXTO EN CLARO

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

Figura 1.23. Tabla de cifrar de Beaufort.

Por lo tanto, se invierte el orden de las letras del alfabeto A y luego se les aplica un desplazamiento hacia la derecha de $(k_i + 1)$ posiciones. Esta afirmación puede comprobarse aplicando la siguiente congruencia:

$$E_{k_i}(A) = (k_i - A) \bmod n = [(n-1) - A + (k_i + 1)] \bmod n \quad 1.19$$

Ejemplo 1.30: Usando la Tabla de Beaufort con clave $K = \text{ULTIMÁTUM}$, cifre el mensaje $M = \text{ESTO ES LA GUERRA SEÑORES}$.

Solución: Siguiendo la tabla de la Figura 1.23 se obtiene:
 $C = \text{QSATI IJUGA HCQMI PHXDH B}$.

Por la operación de sustitución empleada, es posible que, a diferencia del de Vigenère, un carácter se cifre con su valor en claro con una clave distinta de la letra A, como es el caso en el ejemplo anterior en que el segundo (S) y el noveno (G) caracteres se cifran en claro, con las claves (L) y (M), respectivamente. La operación de descifrado, tal como comentábamos, es la misma que la de cifrado, con la excepción que en vez de texto claro M contamos ahora con texto cifrado C, esto es:

$$M_i = D_{k_i}(C_i) = (k_i - C_i) \bmod n \quad 1.20$$

Por ejemplo, para los cinco primeros caracteres del ejemplo anterior (QSATI) aplicando la ecuación 1.20 se obtiene:

$$k_1 = U; C_1 = Q \quad \Rightarrow \quad M_1 = (21-17) \bmod 27 = 4 \quad \Rightarrow \quad M_1 = E$$

$$\begin{array}{llll}
 k_2 = L; C_2 = S & \Rightarrow & M_2 = (11-19) \bmod 27 = 19 & \Rightarrow & M_2 = S \\
 k_3 = T; C_3 = A & \Rightarrow & M_3 = (20-0) \bmod 27 = 20 & \Rightarrow & M_3 = T \\
 k_4 = I; C_4 = T & \Rightarrow & M_4 = (8-20) \bmod 27 = 15 & \Rightarrow & M_4 = O \\
 k_5 = M; C_5 = I & \Rightarrow & M_5 = (12-8) \bmod 27 = 4 & \Rightarrow & M_5 = E
 \end{array}$$

A igual resultado podemos llegar si utilizamos la Tabla de Beaufort para descifrar. De la misma forma que en el método de Vigenère, nos posicionamos en la fila correspondiente al carácter de la clave y buscamos la retícula en la que aparezca el elemento cifrado, su proyección a la fila superior del texto en claro nos entrega el carácter del mensaje. Observe que en este caso, la fila de desplazamiento 0 (letra A) no se corresponde con la del texto en claro como sucedía con Vigenère.

Ejemplo 1.31: Usando la ecuación 1.20 descifre con clave LEIA el siguiente criptograma:
 $C = UKEPL ZÑWTF IHHEG MZOIN H.$

Solución: Siguiendo el método indicado, encontramos el mensaje intergaláctico:
 $M = QUE LA FUERZA TE ACOMPAÑE.$

• Variante del cifrador de Beaufort

Si modificamos la ecuación (1.18) de forma que la función de cifrado E_{k_i} se transforme en $E_{k_i}(M_i) = (M_i - k_i) \bmod n$, el sistema se conoce como variante de Beaufort. Esto es equivalente a cifrar con Vigenère siendo la clave $(n - k_i)$ lo que puede demostrarse a partir de la siguiente congruencia: para un alfabeto A:

$$(A - k_i) \bmod n = [A + (n - k_i)] \bmod n \quad 1.21$$

Al ser el inverso del algoritmo de Vigenère, este cifrador se puede utilizar para descifrar criptogramas obtenidos con Vigenère y viceversa. Algo obvio por lo demás.

1.6.5. Criptoanálisis de los cifrados polialfabéticos periódicos

al utilizar más de un alfabeto, el número de combinaciones de la clave crecerá y también lo hará su entropía y distancia de unicidad. Para un cifrador polialfabético como Vigenère, la distancia de unicidad vendrá dada por el número total de combinaciones usadas para sustituciones simples; esto es si para cada sustitución simple monoalfabeto hay n posibles claves, entonces al utilizar d sustituciones existirán n^d claves posibles.

Ejemplo 1.32: Si el alfabeto de claves son las letras A, B, C y D, a) ¿Cuántas claves de dos elementos podemos formar? b) Encuentre la distancia de unicidad del cifrador de Vigenère para el lenguaje castellano. c) ¿Cuál es su valor para una clave de longitud 10?

Solución: a) Existirán $n^d = 4^2 = 16$ combinaciones posibles. Para este alfabeto de cuatro letras, serán claves: AA, AB, AC, AD, BA, BB, BC, BD, CA, CB, CC, CD, DA, DB, DC y DD.

b) $N = H(K)/D = \log_2(n^d)/D = d * \log_2 n/D = d * \log_2 27/3,4.$

c) Si $d = 10$ entonces $N \approx 10 * 1,4 \approx 14$ caracteres.

Luego, para romper un cifrado polialfabético se necesitará mucho más texto cifrado que en uno monoalfabético, tantas veces como el valor de su período, puesto que en aquel caso la cantidad de texto cifrado necesaria venía dada por $\log_2 n/D$.

• Método de Kasiski

Vamos a profundizar en la característica de periodicidad de los cifradores polialfabéticos. Supongamos que por algún método *aún desconocido por Ud*, logramos adivinar que la longitud de la clave es igual a *seis* caracteres. Si el criptograma C está formado por una cadena de m caracteres $C_1C_2...C_{m-1}C_m$, podemos escribirlo en un formato de *seis* columnas como se indica:

| | | | | | |
|-----------|-----------|-----------|-----------|-----------|----------|
| C_1 | C_2 | C_3 | C_4 | C_5 | C_6 |
| C_7 | C_8 | C_9 | C_{10} | C_{11} | C_{12} |
| C_{13} | C_{14} | C_{15} | C_{16} | C_{17} | C_{18} |
| C_{19} | C_{20} | C_{21} | C_{22} | C_{23} | C_{24} |
| | | | | | |
| C_{m-5} | C_{m-4} | C_{m-3} | C_{m-2} | C_{m-1} | C_m |

Recordando el método de cifrado de Vigenère con una clave de seis caracteres, tenemos que cada una de las seis columnas corresponderá a la cifra con el mismo elemento de la clave. Esto es, los caracteres de una misma columna se corresponden con los de un cifrado monoalfabético por desplazamiento puro dado por el valor del elemento *i*ésimo de clave. Esto nos va a indicar que dos o más caracteres iguales en una columna se deberán a caracteres iguales del texto en claro que, evidentemente, se han cifrado con la misma letra de clave. Además, como el lenguaje castellano presenta una alta redundancia es posible que poligramas característicos tales como *ando*, *ada*, *ido*, *ado*, *ica*, *ita*, *ción*, *mente* y muchos otros sean cifrados con la misma cadena de clave originando cadenas de texto cifrado repetidas en el criptograma.

La probabilidad de que se den estas repeticiones de cadenas será menor que en un cifrador monoalfabético; no obstante, una cadena grande de caracteres repetidos es muy poco probable que aparezca por puro azar. De hecho, trigramas y tetragramas repetidos más de una vez en el criptograma indican una alta probabilidad de que la distancia entre tales cadenas sea un múltiplo de la clave utilizada para cifrar. Aquí está la madre del cordero. Este principio fue observado por el criptólogo alemán *Friedrich W. Kasiski* en 1860 con lo que el método lleva su nombre. En otras palabras, si una clave tiene L caracteres, sólo hay L formas diferentes de posicionar dicha clave sobre la o las palabras en el texto en claro. Esto es, si la clave es NECIO, una repetición típica de cuatro letras como podría ser *ando* (p.e. comprobando, contrabando, bando, bandolero, cantando, abandono, etc.) podrá cifrarse solamente de las siguientes cinco formas ANDO+NECI, ANDO+ECIO, ANDO+CION, ANDO+IONE y ANDO+ONEC.

Luego, para este caso, podemos esperar que un grupo de caracteres que aparecen más de L veces en el texto en claro hayan sido cifrados *al menos* dos veces en la misma posición de la clave y dichas ocurrencias se cifrarán de forma idéntica. Veamos un ejemplo sencillo. Si el mensaje es el famoso monólogo de *Hamlet* y ciframos mod 27 según el método de Vigenère con la clave $K = HAM$ se tiene:

M = T O B E O R N O T T O B E T H A T I S T H E ...
 K = H A M H A M H A M H A M H A M H A M H A M H ...
 C = A O N L O D T O F A O N L T S H T T Z T S L ...

La cadena o secuencia de caracteres *AONL* que aparece dos veces en el criptograma con una separación igual a 9 espacios, sugiere que el período de la clave sea igual a 3 ó 9. Además encontramos la cadena *TS* separada por 6 espacios lo que confirmaría que el período es igual a 3. Con algo más de texto y el uso de las estadísticas del lenguaje como veremos más adelante, seremos capaces de determinar que la clave de este ejemplo tiene efectivamente una longitud de 3 caracteres y que éstos son precisamente *HAM*.

Explicaremos el método de Kasiski mediante un ejemplo detallado. Supongamos que contamos con el criptograma de 404 caracteres que se indica:

PBVRQ VICAD SKAÑS DETSJ PSIED BGGMP SLRPW RÑPWY EDSDE ÑDRDP
 CRCPQ MNPWK UBZVS FNVRD MTIPW UEQVV CBOVN UEDIF QLONM WNUVR
 SEIKA ZYEAC EYEDS ETFFPH LBHGU ÑESOM EHLBX VAEFP UÑELI SEVEF
 WHUNM CLPQP MBRRN BPVIÑ MTIBV VEÑID ANSJA MTJOK MDODS ELPWI
 UFOZM QMVNF OHASE SRJWR SFQCO TWVMB JGRP W 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.

En el criptograma se han subrayado con línea doble algunas cadenas largas que se repiten. Estas son:

- 3 cadenas GGMP: separadas por 256 y 104 caracteres
- 2 cadenas YEDS: separadas por 72 caracteres
- 2 cadenas HASE: separadas por 156 caracteres
- 2 cadenas VSUE: separadas por 32 caracteres.

El valor del máximo común divisor de estas distancias debería ser un múltiplo de la longitud de la clave, esto es: $\text{mcd}(256, 104, 72, 156, 32) = 4$. Luego, la clave podría tener una longitud igual a cuatro caracteres. Hay que tener cuidado con elegir cadenas muy cortas ya que éstas pueden deberse solamente al azar y echar por tierra todas nuestras esperanzas de romper la cifra al aparecer una distancia cuyo valor sea primo con las demás. En nuestro ejemplo sería el caso de elegir, entre otras, también la cadena VR (subrayada) que aparece tres veces al comienzo del criptograma con una separación de 65 y 31 caracteres, que no cumple con el máximo común divisor 4 encontrado anteriormente. Lo mismo ocurre con la cadena RR (subrayada), que se repite dos veces con una separación de 142 y luego 19 espacios. Para más inri, en ambos casos aparece un número primo, lo que asegura que el mcd sea igual a uno.

Si sospechamos que la clave tiene cuatro caracteres, vamos a romper el texto cifrado en cuatro criptogramas independientes C_A , C_B , C_C y C_D de 101 caracteres cada uno y que llamaremos subcriptogramas, tomando para el primero, segundo, tercero y cuarto los caracteres separados por cuatro espacios, siguiendo la escritura en columnas que indicábamos al comienzo de este apartado; es decir:

Primer subcriptograma: $C_1, C_5, C_9, \dots, C_{391}, C_{401}$
 Segundo subcriptograma: $C_2, C_6, C_{10}, \dots, C_{398}, C_{402}$
 Tercer subcriptograma: $C_3, C_7, C_{11}, \dots, C_{399}, C_{403}$
 Cuarto subcriptograma: $C_4, C_8, C_{12}, \dots, C_{400}, C_{404}$

Por lo tanto, tenemos ahora:

$C_A =$ PQAAEPDMRÑEEDCNUSRIECNIONSAAETLUOLAUIEULMNIIEAAOOLU
 MNARSOMRSISERNAISIRTMDTOORLIORRENENOAVSNIAE OFAMTEI
 $C_B =$ BVDÑTSBPPPDÑPPBFBDFPQBUFNUEZCDFBÑMBEÑSFNPBBÑBÑNMKDPF
 QFSJFTBPUNJMBNGDUNUFPFSSÑRPFTPTBTETTJFUBSUTFTPBÑE
 $C_C =$ VISSSIGSWSDCQWZNMWVOEQMVIYESPHEEXEEFWQRP MVISTMSWO
 MOEWQWJWEQEGDISSETEGOSETYWWGQSLGMXOHHECEEIGGIWEE
 $C_D =$ RCKDJEGLRYDRMKVVTVVVDLWRKEYEHGSHVPLVHCPRTVDJJDEIZ
 VHSRCVGVXRUGGLJVEGEGRGTQGVJXGRKRZGUJRRVJHHUEY GKUNU

Para descifrar los caracteres de la clave, suponiendo una cifra del tipo Vigenère, debemos encontrar el desplazamiento que se observa en cada uno de estos cifrados monoalfabéticos. Aplicaremos un *Método por Coincidencia Múltiple* que he llamado *Regla EAOS*. Consiste en observar las frecuencias relativas de los caracteres de cada subcriptograma y marcar las cuatro mayores de forma que sigan modularmente la posición de las letras A, E, O y S en el alfabeto, las cuatro letras más frecuentes del lenguaje castellano. Evidentemente, si se desea una mayor precisión pueden tomarse más letras con el mismo sentido pero en la mayoría de los casos y para lo que aquí nos interesa es suficiente con estas cuatro. Por lo tanto, para estas cuatro letras, los caracteres m que las representen deberán tener una frecuencia relativa alta y estar situadas en las siguientes posiciones:

$m \mod 27$ posición *relativa* de la letra A en el alfabeto desde el origen
 $m+4 \mod 27$ posición *relativa* de la letra E en el alfabeto desde el origen
 $m+15 \mod 27$ posición *relativa* de la letra O en el alfabeto desde el origen
 $m+19 \mod 27$ posición *relativa* de la letra S en el alfabeto desde el origen

Esto quiere decir que al ser todos los subcriptogramas resultado de cifrados monoalfabéticos, entonces alguna letra del texto cifrado tendrá aproximadamente la frecuencia característica de la letra A en el lenguaje, otra la de la letra E, otra de la letra O y otra de la letra S. Estos valores altos de frecuencia deberán estar separados por una relación de distancias constante pues de la A a la E hay cuatro espacios, de la E a la O hay once, de la O a la S cuatro y, por último, de la S a la A siete.

En la Figura 1.24 se muestra la tabla con la contabilización de los caracteres de cada subcriptograma para este 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 |
|-------|----|----|---|---|----|----|----|---|----|---|---|---|---|---|---|----|----|---|----|----|---|---|----|----|---|---|---|
| C_A | 11 | 0 | 2 | 3 | 12 | 1 | 0 | 0 | 11 | 0 | 0 | 5 | 6 | 9 | 1 | 10 | 2 | 1 | 9 | 2 | 4 | 5 | 1 | 0 | 0 | 0 | 0 |
| C_B | 0 | 14 | 1 | 6 | 4 | 12 | 1 | 0 | 0 | 4 | 1 | 0 | 3 | 6 | 8 | 6 | 14 | 2 | 1 | 6 | 9 | 7 | 1 | 0 | 0 | 0 | 1 |
| C_C | 0 | 0 | 1 | 2 | 18 | 0 | 7 | 3 | 2 | 1 | 0 | 1 | 7 | 1 | 0 | 0 | 2 | 6 | 1 | 12 | 3 | 0 | 3 | 12 | 3 | 2 | 1 |
| C_D | 0 | 0 | 3 | 5 | 7 | 0 | 12 | 6 | 1 | 7 | 5 | 4 | 1 | 1 | 0 | 6 | 2 | 1 | 13 | 2 | 3 | 7 | 14 | 0 | 2 | 3 | 2 |

Figura 1.24. Frecuencias de los monogramas del ejemplo de criptoanálisis.

De la Figura 1.24, tomaremos los valores que se subrayan como caracteres en los cuales se cumple en buena medida la regla indicada, es decir:

$$\begin{aligned}(m_A + 4) \bmod 27 &= m_E \\(m_E + 11) \bmod 27 &= m_O \\(m_O + 4) \bmod 27 &= m_S \\(m_S + 7) \bmod 27 &= m_A\end{aligned}$$

1.22

siendo m_A , m_E , m_O y m_S las posiciones de los caracteres con mayor frecuencia relativa en el criptograma que cumplan este sentido modular.

En este caso, siguiendo la Figura 1.24, para el criptograma C_A se observa que la única solución que cumple con la modularidad de (1.22) es aquella en la que el texto cifrado coincide con el texto en claro AEOS con valores (11,12,10,7), luego es posible que la primera letra de la clave sea la A. El valor alto que muestra la letra I no cumple esta condición y se descarta. Para C_B la relación de letras con alta frecuencia en este orden modular está muy claro que se encuentra desplazada un carácter a la derecha, BFPT (14,12,14,9) con lo cual la clave puede ser B. Para el tercer criptograma C_C los números 18, 7, 12, y 12 cumplen con la modularidad exigida por la ecuación obteniendo ahora un ciclo EISW por lo que podemos suponer que la tercera letra de la clave es la E. Por último para el criptograma C_D , elegimos los caracteres RVGK con frecuencias (13,14,12,5) con lo que la clave será la letra R. Con esto llegamos a la conclusión de que la clave puede ser $K = ABER$. Descifraremos entonces el criptograma C según Vigenère con la clave ABER, "a ver" qué sucede.

C = PBVRQ VICAD SKAÑS DETSJ PSIED BGGMP SLRPW RÑPWY
K = ABERA BERAB ERABE RABER ABERA BERAB ERABE RABER
M = PARAQ UELAC OSANO MESOR PREND ACOMO OTROS AÑOSH

Como el texto "Para que la cosa no me sorprenda como otros años..." tiene sentido en castellano, y es imposible que se dé por pura casualidad, seguimos descifrando y obtenemos el texto que se indica. No obstante, observe que la cadena repetida GGMP de criptograma corresponde a la palabra "como", bastante común en nuestro lenguaje. Como ejercicio, encuentre a qué cadenas de caracteres en claro corresponden las cadenas de cifrado YEDS, HASE y VSUE que nos han servido para romper la cifra. El texto completo es:

"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 están ahí, escondidos tras la vida sedentaria y parapetados en fechas como estas de Navidad."

Comienzo del artículo "Gimnasia" del periodista Andrés Aberasturi publicado en el periódico El Mundo el 4/12/94, año del primer curso en que se impartió la asignatura Seguridad Informática en la EUI.

Si los subcriptogramas son relativamente pequeños, a veces no resulta tan fácil decidir cuál es la clave mediante la observación de las frecuencias puesto que será muy aventurado hacer estadísticas con tan pocos datos. En este ejemplo con más de 100 caracteres por cada subcriptograma, las estadísticas no funcionan tan mal como hemos visto. Además, conocido el origen del mensaje, la clave era obvia ¿verdad?

Ejemplo 1.33: *Encuentre la longitud de la clave por el método de Kasiski a partir del criptograma obtenido con el algoritmo de Vigenère.*

C = AWHFW ZAPIE ARXKS CXWEX WEXWE KJPUR EBWTU SCOOB
JGTKJ PUREB WTUSC NGXW.

Solución: *Se observan tres repeticiones XWE seguidas, separadas entre sí por 3 espacios. Además, hay una larga cadena KJPUREBWTUSC que aparece dos veces, separada por 18 espacios. Como esto es muy “sospechoso”, calculamos el mcd $(3, 18) = 3$; luego este valor puede ser la longitud de la clave. De hecho la clave en esta cifra es la cadena PIO. Rompa la cifra en 3 subcriptogramas e intente comprobar si efectivamente se cumplen las estadísticas de frecuencia.*

En el ejemplo anterior, si la clave que se utiliza es *POLLOS* en vez de *PIO*, se obtiene el criptograma C = ADDAD DAWEZ HVXQO XEAE E SZEAE QFLBV EISOB WCVLW PKTQF LBVEI SOBWC TCSD. En él encontramos la cadena de caracteres EAE separada por 6 espacios, además de la cadena QFLBVEISOBWC separada por 18 espacios por lo que la longitud de la clave podría ser $\text{mcd}(6, 18) = 6$. El descifrado con las dos claves y el posterior descubrimiento de esta *tierna canción de infancia* se lo dejo también como ejercicio.

Debe tenerse en cuenta que este método al ser estadístico no es ni mucho menos infalible en el sentido que, por puro azar, se pueden dar cadenas repetidas en el criptograma cuya separación no sea múltiplo de la clave o, incluso peor, que sea un número primo como ya hemos comentado con lo que el máximo común divisor será igual a la unidad. En la práctica, para evitar estas soluciones falsas, debemos contar con un criptograma de muchos caracteres, por ejemplo varias centenas, que incluso una vez roto en subcriptogramas tengamos cerca de la centena de caracteres en cada uno y luego buscar cadenas de caracteres largas, de una longitud mayor o igual a 3 y en lo posible que se repitan más de una vez.

• Índice de Coincidencia

Otra forma de encontrar el período de un cifrador polialfabético es el conocido como *Índice de Coincidencia*, propuesto por William F. Friedman en una publicación de Riverbank en 1922. La publicación de "*El Índice de Coincidencia y sus Aplicaciones en Criptografía*"⁴ considerada durante muchos años como la mayor contribución al desarrollo de la criptología, entronca definitivamente las matemáticas y estadísticas con la criptología y permitió, entre otras cosas, criptoanalizar las máquinas de rotores con

⁴ Friedman, William, "*The Index of Coincidence and Its Applications in Cryptography*", Riverbank Laboratories, publication 22, 1922. También en Khan, David, "*The Codebreakers. The Story of Secret Writing*", Macmillan Publishing Company, New York, 1967, pp. 376-384.

alfabetos progresivos que traían de cabeza a los criptólogos y servicios de inteligencia durante la Segunda Guerra Mundial.

En lo que aquí nos interesa, el Índice de Coincidencia medirá la variación o varianza de la frecuencia de aparición de los caracteres de un criptograma. La idea es la siguiente: en un sistema por sustitución simple monoalfabeto, con período igual a uno, encontramos una importante variación en la frecuencia relativa de aparición de las letras como se desprende de los valores característicos de monogramas de la Figura A.1 del Anexo. En cambio, para los sistemas polialfabéticos con un período grande, la variación de estas frecuencias es muy baja debido al efecto de difusión. Diremos entonces que en el primer caso existe un Índice de Coincidencia IC alto, y que en el segundo este IC será bajo. Definiremos primeramente MD como la *Medida de la Dispersión* que nos entrega la variación de frecuencias en cada carácter, relativa a una *distribución uniforme*:

$$MD = \sum_{i=0}^{n-1} \left(p_i - \frac{1}{n} \right)^2 \quad 1.23$$

donde p_i es la probabilidad de que un carácter cualquiera elegido aleatoriamente del criptograma sea el carácter i ésimo a_i de un alfabeto con longitud n . Además, dado que:

$$\sum_{i=0}^{n-1} p_i = 1 \quad 1.24$$

Si $n = 27$, alfabeto castellano en mayúsculas, se tiene que:

$$MD = \sum_{i=0}^{26} p_i^2 - 0,037 \quad 1.25$$

Ejemplo 1.34: *Demuestre la ecuación (1.25).*

Solución:

$$MD = \sum_{i=0}^{26} p_i^2 - 0,037 = \sum_{i=0}^{26} \left[p_i^2 - \left(\frac{2}{27} \right) p_i + \left(\frac{1}{27} \right)^2 \right]$$

$$MD = \sum_{i=0}^{26} p_i^2 - \frac{2}{27} \sum_{i=0}^{26} p_i + \sum_{i=0}^{26} \left(\frac{1}{27} \right)^2$$

$$MD = \sum_{i=0}^{26} p_i^2 - \frac{2}{27} + 27 \left(\frac{1}{27} \right)^2 \therefore MD = \sum_{i=0}^{26} p_i^2 - 0,037$$

La Medida de Dispersión MD evalúa la altura de los *picos* y los *valles* en una distribución de frecuencias con respecto a una distribución uniforme. Para el lenguaje castellano con $n = 27$, tenemos que $1/n = 0,037$ será la línea base, de forma que los picos serán frecuencias relativas por sobre esta línea y los valles frecuencias relativas por debajo de la misma. Luego, si fr_M es la frecuencia relativa de la letra M, $fr_M - 0,037$ será el tamaño del pico o del valle *observado* y $p_M - 0,037$ el tamaño *esperado* del pico o del valle. Puesto que los picos serán positivos y los valles negativos, para que estos valores no se cancelen en la ecuación de la MD se utiliza $(p_i - 1/n)^2$.

Para una distribución uniforme, esto es las 27 letras del alfabeto equiprobables, se tiene $MD = 27(1/27)^2 - 1/27 = 0$ que es lo esperado pues el Índice de Coincidencia indica la variación de la frecuencia de las letras respecto a una distribución uniforme. Es lógico que un texto que tenga una distribución de caracteres equiprobables presente una medida de dispersión igual a cero. En el otro extremo, si los caracteres del texto presentan la distribución característica del lenguaje castellano se tendrá:

$$\sum_{i=a}^{i=z} p_i^2 = p_a^2 + p_b^2 + p_c^2 + \dots + p_x^2 + p_y^2 + p_z^2 = 0,072 \quad 1.27$$

Luego, la varianza será $MD = 0,072 - 0,037 = 0,035$. Esto quiere decir que la varianza de los caracteres de un criptograma tendrá un valor máximo igual a 0,035 cuando el cifrado haya sido monoalfabético y tiende a cero cuando el cifrado es polialfabético y el número de alfabetos utilizados es muy grande.

$$0 \leq MD \leq 0,035 \quad 1.27$$

El valor de p_i^2 significa la probabilidad de que al tomar dos caracteres aleatorios del criptograma, los dos sean iguales. Este valor se define como Índice de Coincidencia:

$$IC = \sum_{i=0}^{n-1} p_i^2 \quad 1.28$$

Como no conocemos el período ni las probabilidades p_i del criptograma, no será posible encontrar la Medida de la Dispersión, por lo menos de forma teórica. No se preocupe por esto ya que sí será posible, no obstante, *estimar* MD usando la distribución de frecuencias de las letras del texto cifrado y aproximar así la probabilidad con la frecuencia observada. Luego, si f_i son las ocurrencias del carácter i en un criptograma de N letras, la probabilidad de elegir simultáneamente dos caracteres iguales de forma aleatoria, es decir p_i^2 , será:

$$p_i^2 = \frac{f_i(f_i - 1)/2}{N(N - 1)/2} = \frac{f_i(f_i - 1)}{N(N - 1)} \quad 1.29$$

Este resultado se explicará en el siguiente ejemplo. Luego IC será igual a:

$$IC = \frac{\sum_{i=0}^{n-1} f_i(f_i - 1)}{N(N - 1)} \quad 1.30$$

Ejemplo 1.35: *Al cifrar $M = \text{ANIMAL RARO}$ con método de Beaufort y $K = \text{CERDO}$ (no tengo nada contra estos nobles animales) se obtiene $C = \text{CRKRO RNRMA}$.*

- Encuentre la probabilidad de elegir dos caracteres iguales R en C.*
- Demuestre con este criptograma la ecuación (1.29).*

Solución: a) $C = \text{CRKRO RNRMA}$
 01234 56789

Como hay 10 letras en el criptograma y existen 4 caracteres R, el valor de f_R es igual a 4, a lo que podríamos asociar una probabilidad $p_R = 0.4$

b) Tendremos $f_R(f_R-1) = 4(4-1) = 12$ formas de tomar un par de letras R del criptograma (1-3; 1-5; 1-7; 3-1; 3-5; 3-7; 5-1; 5-3; 5-7; 7-1; 7-3 y 7-5) pero como tomar el par 1,3 es igual que tomar el par 3,1 etc., el número efectivo de formas será $f_R(f_R-1)/2$. Por otra parte, como el criptograma tiene 10 letras, habrá $N(N-1) = 10(10-1) = 90$ pares de letras: (0-1; 0-2; 0-3; 0-4; 0-5; 0-6; 0-7; 0-8; 0-9; 1-0; 1-2; 1-3; ... 9-4; 9-5; 9-6; 9-7; 9-8 y 9-9). Pero como el par jk es igual que kj , entonces el número efectivo de pares de letras será $N(N-1)/2$.

Podemos entonces concluir que el Índice de Coincidencia es un método para encontrar de forma aproximada la varianza que presentan los caracteres de un criptograma por medio de la observación de los datos. De la ecuación (1.25) se deduce:

$$IC = MD + 0,037 \quad 1.31$$

En la expresión anterior el valor de IC puede ser calculado a partir de los valores encontrados en un criptograma ya que MD no es posible calcularlo como ya habíamos comentado en un párrafo anterior. No obstante, como el valor de MD se encuentra entre 0 y 0,035 llegamos a la conclusión de que el Índice de Coincidencia IC estará comprendido entre los siguientes valores:

$$0,037 \leq IC \leq 0,072 \quad 1.32$$

Para cifradores con período d , el valor esperado del IC para un texto de N caracteres vendrá dado por la siguiente ecuación:

$$IC = (1/d)[(N-d)/(n-1)] * 0,072 + [(d-1)/d](N/N-1) * 0,037 \quad 1.33$$

| d | IC | d | IC |
|---|-------|--------|-------|
| 1 | 0,072 | 5 | 0,044 |
| 2 | 0,054 | 10 | 0,040 |
| 3 | 0,049 | ... | ... |
| 4 | 0,046 | Grande | 0,037 |

Figura 1.30. Índice de Coincidencia para cifras con período d .

El valor de $IC = 0,072$ para un período d igual a la unidad, esto es un cifrador por desplazamiento puro monoalfabético, nos indica que dicho texto será equivalente al lenguaje castellano en lo que a distribución de frecuencias relativas de los caracteres se refiere. Luego, todo cifrado monoalfabético tendrá este valor que coincidirá con el de un texto en claro, en tanto que aquí no interesa que sea precisamente la letra E la que presente una frecuencia de aparición del 13%; puede ser cualquier otro carácter del criptograma, dependiendo del desplazamiento utilizado en el cifrado.

El Índice de Coincidencia IC presenta una fuerte variación para valores

pequeños del período de cifrado, no así para valores grandes. Por este motivo, en el criptoanálisis de sistemas por sustitución, el método se usa conjuntamente con el de Kasiski puesto que, si bien no es preciso en cuanto al número de alfabetos utilizados, sí lo es para indicar que se trata de una sustitución monoalfabeto o polialfabeto.

Ejemplo 1.36: *Utilizando la ecuación del Índice de Coincidencia, determinar si el siguiente criptograma pertenece a un cifrado por sustitución monoalfabética o polialfabética.*

C = WVKNK BCOFQ NCGEW CEKQO FGNKO FKEGF GEQKO EKFGO EKCFG VGTÑK
OCTUK GNUKI WKGGOV GETKR VQITC ÑCRGT VGOGE GCWOE KHTCF QRQTU
WUVKV WEKQO ÑQOQC NHCDG VKECQ RQKNC NHCDG VKEC

Solución: *Los 139 caracteres del criptograma y su frecuencia son:*

| | | | | | | | | | |
|------|-----|------|-----|------|------|------|------|-----|-----|
| A=0 | B=1 | C=15 | D=2 | E=12 | F=7 | G=16 | H=3 | I=2 | J=0 |
| K=19 | L=0 | M=0 | N=7 | Ñ=3 | O=11 | P=0 | Q=11 | R=4 | S=0 |
| T=7 | U=4 | V=9 | W=6 | X=0 | Y=0 | Z=0 | | | |

Aplicando la ecuación (1.30) se obtiene: $IC = 0,071$. Como este valor se aproxima mucho a $0,072$ se concluye que se trata de un cifrado de tipo monoalfabético. De hecho, el texto en claro es precisamente la cifra del enunciado de este ejemplo "Utilizando la ecuación ... o polialfabética", con un desplazamiento $b = 2$.

No siempre será posible encontrar el período usando el Índice de Coincidencia. Cuando la clave es relativamente larga, más de 4 caracteres, será mucho más confiable el método de Kasiski. De todas maneras, el Índice de Coincidencia nos permitirá determinar, una vez fraccionado el criptograma en subcriptogramas por Kasiski, si cada uno de ellos se trata de un cifrado monoalfabético, comparando el valor encontrado del IC en cada uno de ellos con el característico del lenguaje castellano. Retomemos el ejemplo del artículo "Gimnasia" de nuestro amigo Aberasturi. Para cada uno de los cuatro subcriptogramas se tienen los siguientes valores de IC:

| | | |
|-------|---------------|--------------|
| C_A | \Rightarrow | $IC = 0,070$ |
| C_B | \Rightarrow | $IC = 0,073$ |
| C_C | \Rightarrow | $IC = 0,075$ |
| C_D | \Rightarrow | $IC = 0,065$ |

Como todos los valores se acercan bastante al IC característico del lenguaje castellano ($IC = 0,072$), podemos asegurar que efectivamente cada uno de los cuatro criptogramas hallados se trata de un cifrado monoalfabético como era el caso. En resumen, si se desea atacar un criptograma que suponemos se ha obtenido mediante sustitución con más de un alfabeto, usaremos las siguientes herramientas:

- Análisis de la distribución de frecuencia del texto del criptograma. Si es parecida a la del lenguaje, el cifrado será monoalfabético; caso contrario, será polialfabético.
- Cálculo del Índice de Coincidencia IC para confirmar que el criptograma se debe a un cifrado polialfabético y tener una primera idea del período del cifrador.
- Aplicación del método de Kasiski para encontrar el período, obteniendo varios subcriptogramas. Cálculo luego del IC de cada uno de dichos criptogramas para asegurarse que se trata de un cifrado con un desplazamiento constante.

- d) Uso del *Método de Coincidencia Múltiple* o Regla EAOS para encontrar las letras que forman la clave.
- e) Encontrada la clave, procedemos a descifrar el criptograma siempre en el supuesto de que conocemos el algoritmo de cifra.

1.6.6. Cifradores polialfabéticos no periódicos

La debilidad de los cifradores por sustitución con más de un alfabeto está en la periodicidad de la clave. Esto provoca posibles cadenas repetidas en el criptograma que entrega una pista al criptoanalista, facilitando sobremanera el ataque a estos cifrados. Como la fortaleza del cifrado o distancia de unicidad dependía de la longitud de la clave o período, la solución consistirá pues en aumentar la longitud de esta clave.

¿Qué sucede si aumentamos la longitud de la clave de forma que tenga un tamaño igual o mayor que el texto en claro? Esto sería lo mismo que adoptar el criterio propuesto por Shannon para un sistema con *secreto perfecto* comentado en un capítulo anterior. En este caso los criptogramas soportarían el ataque por el método de Kasiski puesto que al no haber período alguno, sería imposible dividir el criptograma en otros menores.

• Cifrador con clave continua

Si aceptamos que se aumenta la fortaleza de un cifrado por sustitución utilizando una clave tan larga como el mensaje, el problema está ahora en determinar cómo se genera una clave con tales características. Una solución sencilla podría ser elegir como clave un texto, conocido por el transmisor y el receptor claro está, con una cantidad de caracteres a lo mínimo igual que la del mensaje en claro. Por lo tanto, ya no hablamos de una clave sino de una *secuencia de clave* y el cifrador en cuestión dejará de cifrar *bloques* con una clave periódica para convertirse en un cifrador de *flujo* propiamente dicho. Si el método de cifrado es similar al de Vigenère, este criptosistema se conoce como cifrador con clave continua.

Ejemplo 1.37: *Cifre el mensaje que se indica con el algoritmo de Vigenère, utilizando como secuencia de clave el párrafo primero del libro "Cien años de soledad".*

M = INFORMAMOS NEGATIVAMENTE LA COMPRA DE ACCIONES

Solución: *Escribimos el mensaje y la clave conjuntamente y procedemos a cifrar cada par M_i/k_i con la tabla de Vigenère.*

M = INFORMAMOSNEGATIVAMENTELA COMPRADEACCIONES
K = MUCHOSAÑOSDESPUESFRENTEALPELOTONDEFUSILAM
C = THHVGEAZDLPIYPÑMÑFDIZNILLRSWELOPHEHWAXEE

La operación de descifrado de los cifradores de clave continua es obvia. Basta aplicar la operación inversa del desplazamiento iésimo al igual que en los cifradores polialfabéticos, que para el caso de Vigenère (ecuación 1.17) era $M_i = (C_i - k_i) \bmod n$.

Ejemplo 1.38: *Descifre el criptograma cifrado con Vigenère conociendo que la secuencia*

de la clave es el texto de García Lorca que se indica.

$C =$ GSJFE OPEEO UCUGC EIWGP OVVUR WXPPN MRZOL HEMJO
PSIEY PLUGZ LWHCS GETNL C.

$K =$ VERDE QUE TE QUIERO VERDE, VERDE VIENTO, VERDES
RAMAS. EL BARCO SOBRE LA MAR Y EL CABALLO EN LA
MONTAÑA.

Solución:

Realizando la operación de descifrado entre C_i y k_i se obtiene:

$C =$ GSJFE OPEEO UCUGC EIWGP OVVUR WXPPN MRZOL HEMJO PSIEY PLUGZ LWHCS GETNL C
 $K =$ VERDE QUETE QUIER OVERD EVERD EVIEN TOVER DESRA MASEL BARCO SOBRE LAMAR Y
 $M =$ LORCA YVALL EINCL ANSON LASDO SCIMA SDELT EATRO ESPAÑ OLDEL SIGLO VEINT E

El mensaje original es por lo tanto:

$M =$ Lorca y Valle Inclán son las dos cimas del teatro español del siglo
veinte.

1.6.7. Criptoanálisis de los cifrados con clave continua

A pesar de que el espacio de claves es tan grande o más que el de los mensajes, los cifradores con clave continua vistos en el apartado anterior no nos entregan el *ansiado* secreto perfecto; paciencia, ya llegará un sistema con tales características. La razón es que tanto el texto en claro como el texto de la clave presentarán la redundancia característica del lenguaje castellano.

William Friedman propone un método que lleva su nombre y que consiste, básicamente, en observar que una importante cantidad de pares de letras del mensaje en claro y de la secuencia de clave caen dentro de lo que hemos considerado monogramas de *alta frecuencia* del lenguaje. Esto es, existirán pares $M_i k_i$ en los que tanto el carácter M_i del texto en claro como el carácter k_i de la clave tienen ambos una alta frecuencia. Esto reducirá muchísimo el trabajo del criptoanálisis, si lo comparamos con la fuerza bruta de hacer coincidir a cada letra del criptograma todas y cada una de las letras del alfabeto como posibles claves, lo que significaría por ejemplo para los cinco primeros caracteres del criptograma evaluar $27^5 = 14.348.907$ emparejamientos.

Friedman recomienda suponer inicialmente que todos los caracteres del criptograma se corresponden con pares $M_i k_i$ de alta frecuencia. Luego, para cada letra C_i del texto cifrado se escribe como texto en claro el propio alfabeto y como letra clave aquella que da origen al elemento C_i en cuestión. Como veremos a continuación, la distribución de la clave sobre el alfabeto en claro seguirá la fórmula del cifrador de Beaufort. Hecho esto, se procede a buscar los pares M_i y k_i de alta frecuencia que permitirán ir descubriendo simultáneamente el texto en claro y la secuencia de clave. Por simplicidad, supondremos que los caracteres de alta frecuencia se corresponden con los nueve que forman la palabra *ESTIRANDO*. Veamos cómo funciona el método con un texto elegido a propósito para que las cosas nos salgan bien a la primera (jugaré de momento haciendo trampa) con el siguiente par mensaje/clave:

$M =$ SI ESTÁS CANSADO Y HARTO DE TODO, DESCANSA
 $K =$ ANTES DE TIRARTE AL RÍO, PIÉNSALO DOS VECES

Si ciframos el texto con dicha clave, se obtiene:

↓↓↓↓↓↓ ↓↓ ↓↓ ↓↓↓↓ ↓↓↓ ↓↓↓↓ ↓↓↓ ↓↓
 M = SIEST ASCAN SADOY HARTO DETOD ODESC ANSA
 K = ANTES DETIR ARTEA LRIOP IENSA LODOS VECES
 C = SUXWM DWVIE SRWSY RRZIE LIGHD ZRHHU VQUE

Observe que, debido al tipo de mensaje y clave elegidos, casi todos los pares de letras de *TextoEnClaro* y *TextoCifrado* que generan C son de alta frecuencia. Están marcados con una flecha y se indican a continuación.

| | | | | | | | |
|----------------|-------------------------------|----------------|-------------------------------|----------------|-------------------------------|----------------|-------------------------------|
| C _i | M _i K _i | C _i | M _i K _i | C _i | M _i K _i | C _i | M _i K _i |
| S | SA | W | SE | Z | RI | D | DA |
| U | IN | I | AI | I | TO | R | DO |
| X | ET | E | NR | L | DI | H | ED |
| W | SE | R | AR | I | EE | H | SO |
| M | TS | W | DT | G | TN | Q | NE |
| D | AD | S | OE | H | OS | E | AE |

Precisamente esto es lo que nos permitirá romper la cifra. Existirán 27 posibles emparejamientos de letras del texto en claro/clave para cada letra del criptograma. Consideremos las cinco primeras letras del criptograma anterior (*SUXWM*), ordenemos los alfabetos de mensaje y clave de forma que se obtenga siempre S, U, X, W y M, y luego tomemos los pares de alta frecuencia y veamos qué sucede.

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | ↓ | 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 |
| Letra clave: | | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | W | V | U | T | |
| Letra criptograma: | | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | | 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 |
| Letra clave: | | U | T | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | W | V | | | | |
| Letra criptograma: | | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | | 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 |
| Letra clave: | | X | W | V | U | T | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | | |
| Letra criptograma: | | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | X | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | | 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 |
| Letra clave: | | W | V | U | T | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | | | | | | |
| Letra criptograma: | | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | W | | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|--|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | | 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 |
| Letra clave: | | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | W | V | U | T | S | R | Q | P | O | Ñ | N | | | | | | |
| Letra criptograma: | | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | M | | |

| | | | | | |
|--------------|-------|-------|-------|-------|-------|
| Criptograma: | S | U | X | W | M |
| | A - S | D - R | E - T | D - T | E - I |
| | E - O | I - N | T - E | E - S | I - E |
| | O - E | N - I | | I - O | S - T |
| | S - A | R - D | | O - I | T - S |
| | | | | S - E | |
| | | | | T - D | |

Los anteriores son pares de alta frecuencia *LetraEnClaro* - *LetraDeClave* para

estas cinco primeras letras del criptograma. Existirán $4 \times 4 \times 2 \times 6 \times 4 = 768$ posibles pentagramas de texto en claro con secuencia de clave, mucho menos que los 14 millones anteriores. Uno de ellos será el texto en claro *SIEST* con la clave *ANTES*, precisamente los pares 4º (S-A), 2º (I-N), 1º (E-T), 5º (S-E) y 4º (T-S) que se han subrayado en la tabla anterior.

Mientras menor sea esta ventana habrá menor probabilidad de hallar la letra verdadera pero, por otro lado, si consideramos todos los caracteres, el número de combinaciones puede volverse intratable. Con los nueve caracteres de la palabra *ESTIRANDO* se obtienen para cada letra del alfabeto al menos un par M_iK_i de alta frecuencia, excepto cuando el elemento del criptograma es la letra Y. A continuación veremos un ejemplo en el que no resulta afortunada la elección del bloque para comenzar el ataque a un cifrado continuo. Sean el mensaje M y la clave K:

M = SE SUPONE QUE APARECEN DIGRAMAS FRECUENTES EN LOS CIFRADOS
K = CUANDO USAMOS COMO CLAVE UN TEXTO QUE TAMBIÉN ES REDUNDANTE

| | | | | | | | | | | |
|-----|-------|-------|-------|-------|-------|-------|-------|---------|-------|-----------|
| | ↓ | ↓ ↓ | ↓ ↓ ↓ | ↓ ↓ | ↓ ↓ ↓ | ↓ | ↓ | ↓ ↓ ↓ ↓ | ↓ ↓ | ↓ ↓ ↓ ↓ ↓ |
| M = | SESUP | ONEQU | EAPAR | ECEND | IGRAM | ASFRE | CUENT | ESENL | OSCIF | RADOS |
| K = | CUAND | OUSAM | OSCOM | OCLAV | EUNTE | XTOQU | ETAMB | IENES | REDUN | DANTE |
| C = | UYSHS | DHWQG | SSROD | SEONY | MAETP | XMTIY | GÑEYU | MWQQD | GWFCR | UAPIW |

Si consideramos las tres primeras letras del criptograma anterior (UYS) y tomamos los pares de alta frecuencia, obtenemos ahora:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | 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 |
| Letra clave: | U | T | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | W | V | | | | |
| Letra criptograma: | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | U | |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | 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 |
| Letra clave: | Y | X | W | V | U | T | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z |
| Letra criptograma: | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y | Y |

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--------------------|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Letra en claro: | ↓ | 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 |
| Letra clave: | S | R | Q | P | O | Ñ | N | M | L | K | J | I | H | G | F | E | D | C | B | A | Z | Y | X | W | V | U | T | | | | | |
| Letra criptograma: | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | |

Los pares de alta frecuencia LetraTextoClaro - LetraSecuenciaClave serán:

| | | | |
|-------------|-------|---|-------|
| Criptograma | U | Y | S |
| | D - R | - | A - S |
| | I - N | - | E - O |
| | N - I | - | O - E |
| | R - D | - | S - A |

En este caso la elección del comienzo del criptograma en donde aparece la letra Y que es la única que no presenta pares de alta frecuencia con *ESTIRANDO*, nos ha llevado a un *callejón sin salida*. No hay de qué preocuparse, en este tipo de cifra nos será de mucha utilidad la redundancia del lenguaje. Al elegir los 9 caracteres de la palabra *ESTIRANDO* como los de alta frecuencia del lenguaje, se cumple esta condición en pares M_iK_i para aproximadamente el 50% del texto por lo que con un poco

de imaginación llegaremos a describir el mensaje.

Podemos entonces concluir que si se desea cifrar un texto en claro mediante un cifrador con clave continua y que éste tenga un secreto perfecto, deberemos elegir una secuencia de clave aleatoria como sería, por ejemplo, un listado de nombres o direcciones de una guía de teléfonos a partir de una página determinada; si esa guía es antigua y más aún de otro país tanto mejor. La segunda condición que debería cumplir el sistema es que la clave sea única por lo que, tras ser utilizada, debería destruirse. Este principio da lugar a los denominados criptosistemas de uso o control único también denominados genéricamente *one-time pad* comunes hoy en día.

1.6.8. Cifrador de Vernam

En 1917 *Gilbert S. Vernam*, nativo de Brooklyn e ingeniero del MIT que trabaja en los laboratorios de la empresa AT&T, diseña un dispositivo criptográfico para comunicaciones telegráficas basado en los 32 códigos Baudot de los teletipos desarrollados por su compañía. Los códigos Baudot representan los caracteres del lenguaje con cinco elementos que pueden ser el espacio o la marca (el cero y el uno) diseñado para transmisiones telegráficas. Este cifrador, que tuvo una gran aplicación durante la Primera Guerra Mundial, basa su seguridad en el secreto de un clave aleatoria que se supone tan larga como el mensaje y que luego de usarse debería destruirse. Un dato anecdótico: en dicha confrontación algunos encargados del sistema de cifra (llamados *criptocustodios* en el lenguaje militar) hicieron caso omiso de esta recomendación y los códigos fueron rotos por los aliados.

Cada carácter M_i se representa con 5 bits en código Baudot (ver el anexo ya comentado) que se suma OR exclusivo (módulo 2) con la correspondiente clave k_i de una secuencia binaria aleatoria. De esta forma, el cifrador de Vernam genera un flujo de bits de texto cifrado de la forma:

$$C = E_K(M) = C_1 C_2 C_3 \dots C_N \quad 1.34$$

donde:

$$C_i = (M_i + k_i) \bmod 2 \quad \text{para } i = 1, 2, \dots, N \quad 1.35$$

$$C_i = M_i \oplus k_i \quad 1.36$$

Para la operación de descifrado, utilizamos el mismo algoritmo por la propiedad involutiva de la operación OR exclusivo. Esto es:

$$C_i \oplus k_i = (M_i \oplus k_i) \oplus k_i$$

Como $k_i \oplus k_i = 0$ para $k_i = 0$ y $k_i = 1$, se obtiene:

$$C_i \oplus k_i = M_i \quad 1.37$$

En la Figura 1.31 se muestra un cifrador de Vernam binario como el descrito.

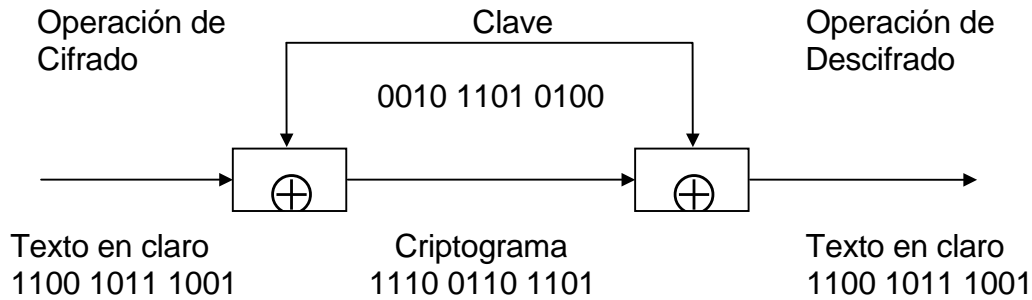


Figura 1.31. Esquema de un cifrador de Vernam binario.

Ejemplo 1.39: Usando el código de Baudot que se encuentra en la tabla del Anexo, cifre el mensaje $M = \text{BYTES}$ con la clave $K = \text{VERNAM}$.

Solución: Haciendo la suma OR exclusivo tenemos:

$$B \oplus V = 11001 \oplus 11110 = 00111 = U$$

$$Y \oplus E = 10101 \oplus 00001 = 10100 = H$$

$$T \oplus R = 10000 \oplus 01010 = 11010 = G$$

$$E \oplus N = 00001 \oplus 01100 = 01101 = F$$

$$S \oplus A = 00101 \oplus 00011 = 00110 = I$$

Luego, $C = \text{UHGF}$

Para descifrar como ya se ha dicho, sencillamente se aplica nuevamente la operación del or exclusivo como se indica en el siguiente ejemplo.

Ejemplo 1.40: Se recibe el siguiente criptograma $C = 00110 \ 10100 \ 11100 \ 11010 \ 00000 \ 00010 \ 01110 \ 01011 \ 00110$ de un texto con código Baudot que se ha cifrado con clave la $K = \text{Gloria Estefan}$. Descífrelo.

Solución: Haciendo la suma OR exclusivo entre C y K ($C \oplus K$) tenemos:

$$K=G: 00110 \oplus 11010 = 11100 = M$$

$$K=L: 10100 \oplus 10010 = 00110 = I$$

$$K=O: 11100 \oplus 11000 = 00100 =$$

$$K=R: 11010 \oplus 01010 = 10000 = T$$

$$K=I: 00000 \oplus 00110 = 00110 = I$$

$$K=A: 00010 \oplus 00011 = 00001 = E$$

$$K=_: 01110 \oplus 00100 = 01010 = R$$

$$K=E: 01011 \oplus 00001 = 01010 = R$$

$$K=S: 00110 \oplus 00101 = 00011 = A$$

Luego $M = \text{MI TIERRA}$.

Como un divertimento más, podemos representar un cifrador de Vernam orientado a caracteres. En este caso la operación de cifra se realiza a través de desplazamientos módulo 27, como si se tratase de un cifrador monoalfabético, con una secuencia de clave compuesta por números aleatorios $NA = k_i$, como se indica.

| | | | | | | | | | | | | | | | | | |
|-------------|---|----|----|----|----|----|----|-----|----|----|----|----|----|----|----|----|----|
| M | = | C | I | F | R | A | D | O | R | D | E | V | E | R | N | A | M |
| M_i | = | 2 | 8 | 5 | 18 | 0 | 3 | 15 | 18 | 3 | 4 | 22 | 4 | 18 | 13 | 0 | 12 |
| k_i | = | 73 | 12 | 39 | 81 | 07 | 28 | 95 | 52 | 30 | 18 | 32 | 29 | 47 | 20 | 07 | 62 |
| $M_i + k_i$ | = | 75 | 20 | 44 | 99 | 7 | 31 | 110 | 70 | 33 | 22 | 54 | 33 | 65 | 33 | 7 | 74 |
| C | = | U | T | Q | R | H | E | C | P | G | V | A | G | L | G | H | T |

Figura 1.32. Cifrado de Vernam orientado a caracteres.

Los valores utilizados para la secuencia de clave en el ejemplo anterior están comprendidos entre 00 y 99, si bien pueden reducirse mod 27 y, por tanto, trabajar sólo con el CCR(27). Del ejemplo anterior podemos destacar un aspecto interesante de un cifrador de Vernam: en el criptograma aparecen caracteres iguales que provienen del

cifrado de caracteres distintos del texto en claro, al igual que en todos los sistemas polialfabéticos, como es el caso de las letras D, E y N que se cifran como el elemento G. Ahora bien, además de utilizar los 27 posibles alfabetos dependiendo del valor de la clave, al ser ésta aleatoria y carecer de periodicidad alguna, hace imposible cualquier tipo de ataque conociendo únicamente el criptograma. Si la secuencia aleatoria de clave usada luego se destruye, entonces el secreto es perfecto. El problema que persiste en este esquema es la transmisión segura de la clave secreta no inventada todavía; para ello debemos esperar hasta el año 1977 en que se presenta el sistema de cifra de clave pública y se soluciona con este método el problema del intercambio de clave.

1.7. CIFRADORES POR SUSTITUCIÓN POLIGRÁMICA MONOALFABETO

Los cifradores poligrámicos, a diferencia de los monográficos que cifraban carácter a carácter, consideran un poligrama con $n \geq 2$ del texto en claro para proceder a su cifrado. De esta forma, el bloque de información a cifrar serán digramas, trigramas o, en general, poligramas. El objeto de este cifrado por sustitución es destruir la distribución de frecuencia típica de los monogramas que, al hacer coincidir el carácter M_i con el elemento cifrado C_i , posibilita el ataque por inspección de frecuencias en el caso de los monoalfabéticos o bien el criptoanálisis mediante método de Kasiski, Índice de Coincidencia y el método de Friedman para los polialfabéticos.

Si suponemos que el cifrador transforma los digramas $M_i M_{i+1}$ del texto en claro en criptogramas $C_i C_{i+1}$, se tendrá que:

$$M = M_1 M_2 \cup M_3 M_4 \cup \dots \cup M_{N-1} M_N \quad 1.38$$

en donde el signo \cup indica *unión* de caracteres y N es el número total de caracteres del mensaje. Luego:

$$E_K(M) = E_K(M_1 M_2) \cup E_K(M_3 M_4) \cup \dots \cup E_K(M_{N-1} M_N) \quad 1.39$$

$$E_K(M) = C_1 C_2 \cup C_3 C_4 \cup \dots \cup C_{N-1} C_N \quad 1.40$$

De los cifradores poligrámicos, los más conocidos son los de Playfair de mediados del siglo XIX, que hace uso de una tabla de cifrar similar a la de Polybios, y el de Hill, que data de comienzos del siglo XX, y que tiene una importancia especial en la criptografía clásica por el hecho de utilizar la matemática de matrices en módulo n para las operaciones de cifrado y descifrado. A comienzos del año 1999, la posibilidad del uso de matrices en los sistemas de cifra volvió a ser considerado, tras la aparición de un algoritmo de clave pública propuesto por una joven irlandesa aunque no tiene nada que ver con el sistema de Hill. No obstante, al final se demostró que el tal invento no era tan espectacular como se suponía.

1.7.1. Cifrador de Playfair

El cifrador de Playfair en realidad fue inventado por *Charles Wheatstone* para comunicaciones telegráficas secretas en el año 1854. No obstante, se le atribuye a su amigo el científico *Lyon Playfair* quien lo presenta a las autoridades inglesas de la época. Nuevamente eso se llama ser un buen amigo. Utilizado por el Reino Unido en la Primera Guerra Mundial, este sistema consiste en separar el texto en claro en digramas y proceder a su cifra de acuerdo a una matriz alfabética de dimensiones 5x5 en la cual se encuentran representadas 25 de las 26 letras del alfabeto inglés.

Para que este método de cifra presente un mayor nivel de seguridad, se incluirá al comienzo de dicha matriz una clave que se escribe a partir de la primera fila omitiendo las letras repetidas. A continuación de dicha clave, se distribuyen las restantes letras del alfabeto hasta completar toda la matriz. Por ejemplo, si la clave es *VERANO AZUL* (*memorable serie de TV de Antonio Mercero*), la matriz será:

En las Figuras 1.33 y 1.34 se muestran las matrices comentadas para el lenguaje castellano de 27 caracteres. En este caso supondremos que las letras I y J ocupan una única celda, al igual que la Ñ y la N. La segunda matriz lleva como clave *VERANO AZUL*, memorable serie de televisión española.

| | | | | |
|---|---|-----|-----|---|
| 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 |

Figura 1.33. Matriz de cifra de Playfair.

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

Figura 1.34. Matriz de cifrado de Playfair con clave *VERANO AZUL*.

El método de Playfair cifrará pares de caracteres del texto en claro M_1M_2 como C_1C_2 de acuerdo a las siguientes reglas:

- Si M_1 y M_2 se encuentran en la misma fila, se eligen los elementos del criptograma C_1 y C_2 como aquellos que están a la derecha de M_1 y M_2 , respectivamente. Esta operación se realiza módulo 5, de forma que para la matriz de la Figura 1.34 se cumplen las siguientes transformaciones:

Pares del texto en claro

Criptograma

| | |
|--------------|----|
| EA (1ª fila) | RN |
| LU (2ª fila) | BL |
| DH (3ª fila) | FC |

- b) Si M_1 y M_2 se encuentran en la misma columna, se eligen C_1 y C_2 como los caracteres que están inmediatamente debajo de ellos, operando módulo 5. Para la matriz de la Figura 1.34 se cumplen las siguientes transformaciones:

| Pares del texto en claro | Criptograma |
|--------------------------|-------------|
| ED (2ª columna) | ZK |
| FU (3ª columna) | MF |
| AX (4ª columna) | LA |

- c) Si M_1 y M_2 se encuentran en filas y columnas distintas, entonces forman dos vértices de un rectángulo. Los elementos C_1 y C_2 se obtienen de los dos vértices que faltan para completar dicha figura geométrica, considerando siempre la fila de M_1 como el elemento C_1 . Esto es, en la matriz de la Figura 1.34 se cumplen las siguientes transformaciones:

| Pares del texto en claro | Criptograma |
|--------------------------|-------------|
| OT | ZS |
| YU | WB |

Recuerde, además, que las letras I y J ocupan una misma celda, al igual que la N y la Ñ, por lo que se cumplen por ejemplo también las siguientes transformaciones en dicha matriz:

| Pares del texto en claro | Criptograma |
|--------------------------|-------------|
| MI = MJ | PK |
| EN = EÑ | RV |

No obstante, si en la operación de descifrado caemos en la retícula I/J, siempre descifraremos como la letra I.

- d) Al representar el texto en claro como una cadena de digramas, pueden aparecer caracteres repetidos con lo cual no podría aplicarse ninguna de las tres opciones de cifrado anteriores. Tal sería el caso de cifrar el siguiente mensaje tenebroso:

M = LAS SOMBRAS LLAMAN A LA PUERTA DEL CASTILLO.
 M = LA SS OM BR AS LL AM AN AL AP UE RT AD EL CA ST IL LO.

La solución a este problema (los digramas segundo SS y sexto LL) está en romper esta repetición antes del cifrado, incluyendo una letra nula que, de acuerdo al lenguaje castellano podría ser X, Z o Q por ejemplo. Usaremos en el texto la letra X como carácter de relleno, con lo que el mensaje se transforma ahora en:

M = LA SX SO MB RA SL LA MA NA LA PU ER TA DE LC AS TI LX LO.

Observe que al incluir este carácter de relleno, ha desaparecido la repetición LL del digrama sexto, si bien aparece otro nuevo (LL) ahora en la posición 18 que se rompe de igual manera añadiendo la letra X.

- e) Por último, es posible que el mensaje final a cifrar, una vez eliminados los digramas repetidos, tenga un número impar de caracteres. En tal situación se añade un carácter nulo al final de la cadena para poder cifrar el último carácter del texto en claro. Por ejemplo, si al mensaje anterior le añadimos la palabra *HOY*, además de ser más inquietante, quedaría como sigue:

M = LA SX SO MB RA SL LA MA NA LA PU ER TA DE LC AS TI LX LO HO YX.

Ejemplo 1.41: *Cifre el mensaje, M = "Las sombras llaman a la puerta del castillo hoy" con una matriz de Playfair con clave K = MIEDO.*

Solución: *Siguiendo la matriz con clave MIEDO, ciframos los digramas del mensaje, a saber: M = LA SX SO MB RA SL LA MA NA LA PU ER TA DE LC AS TI LX LO HO YX, obteniendo:*
C = HCXEUEIA QBXSHCAH HFHCUZIS QFODSLCQ RDSEPEPM ZY.
La matriz de cifra y comprobación de la misma es tarea suya.

Para descifrar un criptograma obtenido mediante Playfair, simplemente utilizamos el algoritmo inverso, esto es:

- Si los elementos C_1 y C_2 están en la misma fila, se eligen M_1 y M_2 como los caracteres inmediatamente a la izquierda, módulo 5.
- Si los elementos C_1 y C_2 se encuentran en la misma columna, se eligen M_1 y M_2 como los caracteres inmediatamente arriba de aquellos, módulo 5.
- Si los elementos C_1 y C_2 están en filas y columnas distintas, M_1 y M_2 se eligen como aquellos caracteres que forman los vértices que faltan del recuadro, comenzando por la fila del primer elemento C_1 .

Ejemplo 1.42: *Si en la matriz de cifra de Playfair con la clave BEATLES se eliminan los caracteres K y Ñ, con relleno X, descifre el siguiente criptograma:*
C = EC TB AZ EN WB JH TX AB BU VC LO JT PM IL.

Solución: *Los digramas descifrados con la matriz que habrá encontrado serán:*
EC \Rightarrow WE; TB \Rightarrow AL; AZ \Rightarrow LX; EN \Rightarrow LI; WB \Rightarrow VE;
JH \Rightarrow IN; TX \Rightarrow AY; AB \Rightarrow EL; BU \Rightarrow LO; VC \Rightarrow WS;
LO \Rightarrow UB; JT \Rightarrow MA; PM \Rightarrow RI; IL \Rightarrow NE.
El texto en claro es M = WE ALL LIVE IN A YELLOW SUBMARINE.

1.7.2. Criptoanálisis del cifrado de Playfair

¿Qué podemos decir acerca de la distancia de unicidad del cifrador de Playfair? A simple vista, el poder ordenar aleatoriamente los 25 caracteres de la matriz, parece que la equivocación de la clave será el factorial de 25. No

obstante, debido al algoritmo de cifra propuesto por Playfair, no todas las matrices de 5x5 son distintas. Vamos a verlo. Considérese la siguiente matriz de cifra con clave *PATOS* y la posterior rotación de las filas tres posiciones hacia abajo y, a continuación, una rotación de las columnas una posición hacia la izquierda:

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

Figura 1.35. Matriz de Playfair con clave *PATOS*.

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

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

Figura 1.36. Matrices recíprocas a la principal de la Figura 1.35.

Estas y otras matrices que se obtengan por rotaciones de filas y/o columnas son recíprocas de la primera puesto que se obtienen los mismos resultados al cifrar un texto según el método de Playfair. Por ejemplo, si ciframos con estas tres matrices el mensaje $M = LA\ PATA\ Y\ EL\ PATO\ TUVIERON\ PATITOS$, compruebe se obtiene el mismo criptograma $C = HSAT\ OTOK\ GSTO\ SOMZ\ KDYE\ MATO\ QDSP$. No obstante, esto sólo reduce el número de matrices posible en $4 \times 4 = 16$ que serán las recíprocas de la principal por lo que el valor de la entropía de la clave se asemeja mucho a la de un sistema del César con clave visto en el apartado 1.4.3.

Por otra parte, el cifrado de Playfair presenta una debilidad que facilita el criptoanálisis. Conociendo qué palabras pueden ser comunes en el texto que se intenta romper, con la ayuda de las estadísticas de digramas comunes del lenguaje, se puede ir confeccionando la matriz y, finalmente, descifrar el criptograma. A continuación se comenta el procedimiento a seguir en este tipo de ataque; el lector interesado en profundizar sobre el tema puede consultar la referencia que se indica.⁵

Sabemos que en el lenguaje existen digramas más comunes que otros y como este sistema cifrará siempre el digrama M_1M_2 como el mismo criptograma C_1C_2 , es lógico esperar una correspondencia de esta redundancia del lenguaje en el texto

⁵ Konheim, Alan G., "*Cryptography: A Primer*", John Wiley & Sons, 1981, pp. 95-110.

cifrado. Siguiendo la tabla de digramas del anexo, encontramos los siguientes pares con una frecuencia relativa superior a 500 para un texto con 40.000 caracteres.

| <u>Digrama</u> | <u>nº veces</u> | <u>Inverso</u> | <u>nº veces</u> |
|----------------|-----------------|----------------|-----------------|
| DE | 1084 | ED | 290 |
| ES | 1010 | SE | 547 |
| EN | 901 | NE | 370 |
| OS | 764 | SO | 212 |
| AD | 649 | DA | 436 |
| TE | 639 | ET | 115 |
| IN | 610 | NI | 191 |
| ER | 563 | RE | 537 |
| AS | 560 | SA | 227 |
| EL | 559 | LE | 245 |
| OR | 544 | RO | 372 |
| NT | 536 | TN | 24 |
| ST | 535 | TS | 19 |
| RA | 520 | AR | 493 |

Figura 1.37. Digramas más frecuentes del lenguaje castellano y sus inversos.

Luego, si seguimos la matriz de cifrado de la Figura 1.35 con la clave PATOS se obtendrán, entre otros, los siguientes pares de cifrados:

| | | | | | | | |
|----------|----------|----------|----------|----------|----------|----------|----------|
| M_1M_2 | C_1C_2 | M_1M_2 | C_1C_2 | M_1M_2 | C_1C_2 | M_1M_2 | C_1C_2 |
| DE | EF | ED | FE | ES | FO | SE | OF |
| EN | CR | NE | RC | ER | KY | RE | YK |

Es decir, en este criptograma se verán representados todos los digramas del texto en claro, sólo que como digramas distintos; esto es, *FO* en vez de *ES*, *YK* en vez de *RE*, etc. Por lo tanto, si observamos que en el texto cifrado aparece, por ejemplo, el digrama *XB* con alta frecuencia y, simultáneamente, el digrama inverso *BX* con muy baja frecuencia, según la tabla de la Figura 1.37 podríamos suponer que se trata del digrama en claro *NT* o bien *ST* que cumplen con esta característica. Si encontramos el digrama *HK* con alta frecuencia del texto cifrado y el inverso *KH* también tiene una frecuencia similar, podría tratarse de los digramas en claro *ER* o *RA*. El digrama más frecuente del criptograma podría ser *DE*, *ES* o *EN* en el texto en claro, etc.

Una vez determinadas algunas correspondencias, el criptoanalista también tiene en cuenta los caracteres que siguen a dichos digramas para formar trigramas y así ir estableciendo las letras equivalentes al alfabeto. Además, si conoce alguna palabra que supuestamente se repite en el texto en claro y que contenga entre sus caracteres las equivalencias anteriores, podrá encontrar más equivalencias y éstas otras más, como si se tratase de un procedimiento en cascada. Un análisis detallado del método está fuera del alcance de este libro; no obstante, siguiendo estas pautas y con *mucha paciencia*, este método a modo de un *entretenido puzzle* nos lleva finalmente a la consecución de la matriz de cifra de Playfair.

1.7.3. Cifrador de Hill

En 1929, *Lester S. Hill*, un joven matemático, publica en Nueva York un artículo en el que propone la utilización del álgebra y, en particular de las matrices, en la operación de cifrado. La importancia del método de cifra propuesto por Hill descansa en la utilización de transformaciones lineales matriciales operando en módulo 26 -las letras del alfabeto inglés- con lo cual se facilita el cifrado poligráfico, algo que tras el invento de Playfair fue insistentemente buscado por los criptólogos y matemáticos de la época. Desgraciadamente para Hill, su invento, aunque muy interesante para los científicos en aquella época, no era fácil de implantarlo en una máquina –no se había inventado el ordenador- y no pudo competir con otros criptógrafos que proliferaron en esos años como fue la máquina Enigma de los alemanes y aparatos de cifra como los de Hagelin.

Dado que bajo ciertas condiciones este sistema presenta una alta seguridad, que puede implementarse fácilmente en los ordenadores actuales y que hace uso de una buena cantidad de conceptos de la aritmética modular operando con matrices, profundizaremos en este cifrador y en su criptoanálisis. A pesar del alto valor de la posible entropía de su clave, vamos a ver que si conocemos el texto en claro y su criptograma asociado, este sistema no soporta un criptoanálisis.

Inicialmente, Hill plantea el problema como el conjunto de cuatro ecuaciones que se indican a continuación, en donde la variable y_i representa las letras cifradas y la variable x_i las letras del texto en claro.

$$\begin{aligned} y_1 &= 8x_1 + 6x_2 + 9x_3 + 5x_4 \pmod{26} \\ y_2 &= 6x_1 + 9x_2 + 5x_3 + 10x_4 \pmod{26} \\ y_3 &= 5x_1 + 8x_2 + 4x_3 + 9x_4 \pmod{26} \\ y_4 &= 10x_1 + 6x_2 + 11x_3 + 4x_4 \pmod{26} \end{aligned} \quad \boxed{1.41}$$

Por otra parte, Hill define un alfabeto de cifrado arbitrario como el que se indica en la Figura 1.38 aunque en el libro usaremos el habitual: A = 0, B = 1, ... Z = 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 |
| 5 | 23 | 2 | 20 | 10 | 15 | 8 | 4 | 18 | 25 | 0 | 16 | 13 | 7 | 3 | 1 | 19 | 6 | 12 | 24 | 21 | 17 | 14 | 22 | 11 | 9 |

Figura 1.38. Alfabeto de cifrado propuesto por Hill.

El mensaje original utilizado por Hill era $M = \text{DELAY OPERATIONS}$. Toma entonces los cuatro primeros elementos, el tetragrama *DELA*, que corresponden a las variables x_1, x_2, x_3 y x_4 , reemplaza sus valores de acuerdo al alfabeto indicado en la ecuación 1.41 para obtener los primeros cuatro elementos del criptograma. Repite esta operación con los tetragramas restantes *YOPE*, *RATI* y *ONS* y con ello obtiene el criptograma $C = \text{JCOW ZLVB DVLE QMXC}$. Como en todo sistema poligrámico, se usan elementos de relleno si el último bloque tiene un tamaño menor. Para descifrar este criptograma, basta con resolver ahora el siguiente sistema de ecuaciones en x :

$$\begin{aligned} x_1 &= 23y_1 + 20y_2 + 5y_3 + 1y_4 \pmod{26} \\ x_2 &= 2y_1 + 11y_2 + 18y_3 + 1y_4 \pmod{26} \\ x_3 &= 2y_1 + 20y_2 + 6y_3 + 25y_4 \pmod{26} \\ x_4 &= 25y_1 + 2y_2 + 22y_3 + 25y_4 \pmod{26} \end{aligned} \quad \boxed{1.42}$$

La comprobación de la cifra completa del mensaje de Hill y su posterior recuperación se lo dejo como ejercicio para más adelante, pero por ahora sigamos con la explicación de este método. De lo anterior, podemos deducir que el cifrado de Hill se trata de un cifrador por sustitución monoalfabética y poligrámico, en tanto que sustituye d caracteres del texto en claro por d caracteres de texto cifrado; en este caso $d = 4$. Si representamos el problema de Hill de cuatro ecuaciones mediante matrices, se tiene que la transformación para la operación de cifra será:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} 8 & 6 & 9 & 5 \\ 6 & 9 & 5 & 10 \\ 5 & 8 & 4 & 9 \\ 10 & 6 & 11 & 4 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} \quad 1.43$$

Para la operación de descifrado se tendrá entonces:

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 23 & 20 & 5 & 1 \\ 2 & 11 & 18 & 1 \\ 2 & 20 & 6 & 25 \\ 25 & 2 & 22 & 25 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} \quad 1.44$$

Las operaciones anteriores podemos generalizarlas suponiendo $\{y\} = C$, $\{x\} = M$ y la matriz $\{K\}$ de orden $d \times d$ como la clave K , luego:

$$C = K_E \times M \text{ mod } n \quad 1.45$$

$$M = K_D \times C \text{ mod } n \quad 1.46$$

En las ecuaciones anteriores, C y M serán vectores columna de dimensiones $d \times 1$, siendo d el tamaño del d -grama. En el caso del mensaje M , corresponderá al bloque de texto en claro de tamaño d -grama que se desea cifrar y para el criptograma C serán los d -grama elementos obtenidos al realizar la multiplicación de $\{K_E\}$ por $\{M\}$ reduciendo los resultados módulo n . Observe que $\{K_D\}$ deberá ser la matriz inversa de la matriz de cifra $\{K_E\}$.

Como se ha comentado, al trabajar con bloques de información igual a d caracteres, es posible que el mensaje M no sea múltiplo de este valor. En tales circunstancias, se rellenará el último bloque hasta completar el d -grama con un carácter nulo que deberá ser conocido por quienes comparten el cifrado; por ejemplo la letra X . En cuanto al alfabeto de cifrado, si bien puede utilizarse cualquiera como lo propuso Hill, le recuerdo que en el libro usaremos la representación numérica habitual en módulo 27, es decir, $A = 0$, $B = 1$, etc.

Consideraciones sobre la matriz K

La matriz K será siempre cuadrada, y sus elementos serán nuestra clave secreta. Será, además, el punto más importante del criptosistema, donde reside su seguridad. No será suficiente el hecho de que la clave sea una matriz cuadrada; ésta deberá cumplir ciertos requisitos que pasamos a enumerar:

- a) Deberá ser una matriz de dimensión $d \times d$, que viene dada por el d -grama que vamos a cifrar. Puesto que en nuestro caso el d -grama a cifrar tiene d filas, la matriz clave deberá tener d columnas para que el producto sea factible. Por otra parte, como queremos que el resultado sea otro d -grama o matriz columna, entonces la matriz clave deberá tener d filas ya que el resultado hereda el número de filas del primer factor de la multiplicación y el número de columnas del segundo factor.
- b) Los elementos de la matriz serán enteros que formen parte del Conjunto Completo de Restos módulo n , en nuestro caso para el lenguaje castellano en mayúsculas $[0, 26]$. Utilizar números fuera de este rango no tiene sentido pues caeríamos en una clase de equivalencia de dicho módulo. Una fórmula interesante de recordar los números que intervienen en la matriz clave, consiste en asignar letras a dichos números; en este caso diremos que dicha matriz con letras se trata de una *matriz simbólica*. La Figura 1.39 muestra un ejemplo de matriz simbólica para cifrado de trigramas.

$$K = \begin{pmatrix} P & E & L \\ I & G & R \\ O & S & O \end{pmatrix} = \begin{pmatrix} 16 & 4 & 11 \\ 8 & 6 & 18 \\ 15 & 19 & 15 \end{pmatrix} \quad \boxed{1.47}$$

Figura 1.39. Matriz simbólica trigrámica con clave PELIGROSO.

La mayor utilidad de la matriz simbólica está en el intercambio de claves entre transmisor y receptor. Recuerde que en aquel entonces no había nacido aún la criptografía de clave pública; además, siempre es más humano recordar una clave como un conjunto de letras o palabras y no como un grupo de números sin sentido.

- c) La matriz K no deberá ser singular, es decir, deberá tener inversa para poder realizar el proceso de descifrado según se indicaba en la ecuación 1.46. Para demostrar que una matriz no es singular, basta con demostrar que el determinante es distinto de cero.

Para encontrar la matriz inversa de K , haremos:

$$K^{-1} = \frac{T_{Adj(K)}}{|K|} \quad \boxed{1.48}$$

donde K^{-1} es la matriz Inversa de K , $T_{Adj(K)}$ es la Traspuesta de la matriz Adjunta de K , y $|K|$ es el determinante de K .

Aunque se supone un conocimiento básico de la aritmética matricial, a continuación explicaremos brevemente las operaciones necesarias para el cálculo de la matriz inversa. Dada una matriz K , su traspuesta $T_{(K)}$ será aquella en la que los elementos (i, j) se intercambian por los elementos (j, i) , es decir se intercambian filas por columnas, como se indica en la siguiente ecuación:

$$\text{Si } K = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \text{ entonces } T_{(K)} = \begin{pmatrix} k_{11} & k_{21} & k_{31} \\ k_{12} & k_{22} & k_{32} \\ k_{13} & k_{23} & k_{33} \end{pmatrix} \quad \boxed{1.49}$$

Ejemplo 1.43: *Encuentre la matriz trigrámica traspuesta de la matriz con clave simbólica $K = \text{NO ESTA MAL}$.*

Solución: $N = 13; O = 15; E = 4; S = 19; T = 20; A = 0; M = 12; A = 0; L = 11$. Así:

$$K = \begin{pmatrix} 13 & 15 & 4 \\ 19 & 20 & 0 \\ 12 & 0 & 11 \end{pmatrix} \text{ luego } T_{(K)} = \begin{pmatrix} 13 & 19 & 12 \\ 15 & 20 & 0 \\ 4 & 0 & 11 \end{pmatrix}$$

Si la matriz anterior sirve o no para cifrar mensajes en castellano módulo 27 (si está o no mal como dice el ejemplo) lo podrá comprobar un poco más adelante. Se llama matriz adjunta de K ($Adj_{(K)}$) a aquella que se obtiene al sustituir cada elemento a_{ij} por su adjunto correspondiente o, lo que es lo mismo, los determinantes de los elementos a_{ij} de la matriz K . Sea $|a_{ij}|$ dicho determinante, entonces:

$$K = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \quad Adj_{(K)} = \begin{pmatrix} |a_{11}| & -|a_{12}| & |a_{13}| \\ -|a_{21}| & |a_{22}| & -|a_{23}| \\ |a_{31}| & -|a_{32}| & |a_{33}| \end{pmatrix} \quad \boxed{1.50}$$

Para obtener $|a_{ij}|$ eliminamos la fila i y la columna j y con los elementos que quedan calculamos su determinante. Por ejemplo, en la matriz de la ecuación (1.50) el elemento $|a_{11}| = k_{22} \cdot k_{33} - k_{32} \cdot k_{23}$; $|a_{22}| = k_{11} \cdot k_{33} - k_{31} \cdot k_{13}$; etc.

Ejemplo 1.44: *Para la matriz con los valores que se indican, se pide encontrar su matriz adjunta: $k_{11}=2, k_{12}=4, k_{13}=6, k_{21}=3, k_{22}=5, k_{23}=7, k_{31}=1, k_{32}=9$ y $k_{33}=0$.*

Solución: La matriz en cuestión es: $K = \begin{pmatrix} 2 & 4 & 6 \\ 3 & 5 & 7 \\ 1 & 9 & 0 \end{pmatrix}$

Luego los valores de los determinantes $|a_{ij}|$ son los siguientes:

$$|a_{11}| = (5 \cdot 0 - 9 \cdot 7) = -63$$

$$|a_{12}| = (3 \cdot 0 - 1 \cdot 7) = -7$$

$$|a_{13}| = (3 \cdot 9 - 1 \cdot 5) = 22$$

$$|a_{21}| = (4 \cdot 0 - 9 \cdot 6) = -54$$

$$|a_{22}| = (2 \cdot 0 - 1 \cdot 6) = -6$$

$$|a_{23}| = (2*9 - 1*4) = 14$$

$$|a_{31}| = (4*7 - 5*6) = -2$$

$$|a_{32}| = (2*7 - 3*6) = -4$$

$$|a_{33}| = (2*5 - 3*4) = -2$$

Por lo tanto, la matriz adjunta de K , $Adj_{(K)}$ será:

$$Adj_{(K)} = \begin{pmatrix} |a_{11}| & -|a_{12}| & |a_{13}| \\ -|a_{21}| & |a_{22}| & -|a_{23}| \\ |a_{31}| & -|a_{32}| & |a_{33}| \end{pmatrix} = \begin{pmatrix} -63 & 7 & 22 \\ 54 & -6 & -14 \\ -2 & 4 & -2 \end{pmatrix}$$

Para ejercitarse un poco con estos cálculos, encuentre la matriz adjunta de la clave simbólica es $K = \text{ESTO NO ESTA TAN MAL}$.

Para poder encontrar la matriz inversa, deberá cumplirse que el determinante de ésta sea distinto de cero; en caso contrario diremos que la matriz es singular y, por tanto, no podrá ser utilizada como clave para cifrar. Ahora bien, puesto que estamos trabajando dentro de un cuerpo, esta condición de singularidad debe darse también dentro de él. Esto significa que el valor del determinante de la matriz clave reducido a módulo n tampoco debe ser igual a cero para que esta sea válida; es decir, deberá cumplirse que $|K| \bmod n \neq 0$ como se indica en el siguiente ejemplo.

Ejemplo 1.45: ¿Pueden utilizarse estas matrices para cifrar un mensaje en módulo 27?

$$K_1 = \begin{pmatrix} 1 & 2 & 2 \\ 1 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix} \quad K_2 = \begin{pmatrix} 12 & 3 \\ 2 & 5 \end{pmatrix}$$

Solución:

El determinante de K_1 para cifrar trigramas será igual a:

$$|K_1| = k_{11}(k_{22}*k_{33} - k_{32}*k_{23}) - k_{12}(k_{21}*k_{33} - k_{31}*k_{23}) + k_{13}(k_{21}*k_{32} - k_{31}*k_{22})$$

$$|K_1| = 1(1*0 - 0*1) - 2(1*0 - 1*1) + 2(1*0 - 1*1) = 0, \text{ luego } K_1 \bmod 27 = 0.$$

La matriz de K_1 es singular y no puede ser usada para cifrar.

El determinante de K_2 para cifrar digramas será igual a:

$$|K_2| = k_{11}*k_{22} - k_{21}*k_{12} = (12*5 - 2*3) = 54$$

$$|K_2| \bmod 27 = 54 \bmod 27 = 0$$

La matriz de K_2 tampoco puede ser usada para cifrar en módulo 27 al ser singular dentro del cuerpo.

Además del requisito $|K| \neq 0$, debemos recordar que en criptografía trabajamos con números comprendidos entre 0 y $n-1$, el cuerpo de la cifra, por lo que no nos servirán los números fraccionarios. La ecuación (1.48) nos indica que para la existencia de la matriz inversa K^{-1} deberá ser posible dividir por el determinante de K ; es decir, debe existir el inverso de $|K|$ en módulo n . Por lo tanto, dicha ecuación podrá escribirse como sigue:

$$K^{-1} = T_{Adj(K)} * \text{inv}(|K|, n) \bmod n \quad \boxed{1.51}$$

en donde se cumplirá que $\text{inv}(|K|, n) * |K| \bmod n = 1$. En realidad, este resultado nos permitirá encontrar la denominada matriz de identidad I que veremos más adelante.

Como bien sabemos, el valor de este inverso no siempre existe; la condición necesaria para su existencia es que $|K|$ y el módulo n sean primos entre sí. Seguro que ya está convencido de esto.

Recapitulando entonces, en el cifrado de Hill deberá existir una transformación inversa que permita recuperar el mensaje en claro, y dicha operación sólo la puede dar la matriz clave. Esto quiere decir que las claves K_E y K_D que aparecían en las ecuaciones (1.45) y (1.46) deberán ser matrices inversas en el módulo de trabajo. Luego, la condición *sine qua non* del cifrador de Hill es que la matriz de cifrado tenga inversa, es decir:

$$K_D = K_E^{-1} \quad \boxed{1.52}$$

Ejemplo 1.46: a) Compruebe que la operación $[K] * [K^{-1}] = I$, en donde I es la matriz de identidad.

b) Compruebe este resultado en particular para la matriz K con elementos $k_{11} = 3$, $k_{12} = 2$, $k_{21} = 1$ y $k_{22} = 4$ en módulo 27.

Solución: a) Si $K = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix}$ entonces $|K| = (k_{11} \times k_{22} - k_{21} \times k_{12})$

y la matriz inversa será:

$$K^{-1} = \frac{1}{|K|} \begin{pmatrix} k_{22} & -k_{12} \\ -k_{21} & k_{11} \end{pmatrix}$$

Si multiplicamos $K^{-1} * K$ se tiene:

$$\frac{1}{|K|} \begin{pmatrix} k_{22} & -k_{12} \\ -k_{21} & k_{11} \end{pmatrix} \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I$$

b) Para la matriz indicada de 2×2 se tiene:

$$K = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} \text{ mod } 27 \quad \text{Luego, } |K| = (3 \times 4 - 1 \times 2) \text{ mod } 27 = 10$$

$$K^{-1} = \frac{1}{10} \begin{pmatrix} 4 & -2 \\ -1 & 3 \end{pmatrix} \text{ mod } 27$$

Como $\text{inv}(10, 27) = 19$ entonces:

$$K^{-1} = \begin{pmatrix} 4 \times 19 & -2 \times 19 \\ -1 \times 19 & 3 \times 19 \end{pmatrix} \text{ mod } 27 = \begin{pmatrix} 22 & 16 \\ 8 & 3 \end{pmatrix} \text{ mod } 27$$

Luego, multiplicando $K * K^{-1}$:

$$K * K^{-1} = \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} \begin{pmatrix} 22 & 16 \\ 8 & 3 \end{pmatrix} \text{ mod } 27 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ mod } 27$$

Los valores del producto de las matrices son:

$$l_{11} = (3 \times 22 + 2 \times 8) \text{ mod } 27 = (12 + 16) \text{ mod } 27 = 1$$

$$l_{12} = (3 \times 16 + 2 \times 3) \text{ mod } 27 = (21 + 6) \text{ mod } 27 = 0$$

$$l_{21} = (1 \times 22 + 4 \times 8) \text{ mod } 27 = (22 + 5) \text{ mod } 27 = 0$$

$$l_{22} = (1 \times 16 + 4 \times 3) \text{ mod } 27 = (16 + 12) \text{ mod } 27 = 1$$

Por último, al igual que en todos los sistemas que implican una multiplicación en el cifrado, para que exista la operación inversa debe cumplirse que el determinante de

la matriz clave no sea múltiplo con el orden del grupo en el que se trabaja. Sólo si $\text{mcd}[|K|, n] = 1$ se podrá usar esa matriz K para cifrar. Luego, la condición suficiente y necesaria de esta matriz K será:

$$|K| \bmod n \neq 0 \quad 1.53$$

$$\text{mcd}[|K|, n] = 1 \quad 1.54$$

Ejemplo 1.47: ¿Se puede utilizar esta matriz K para cifrar digramas en módulo 27 cuyos elementos son $k_{11} = 4$, $k_{12} = 3$, $k_{21} = 2$, $k_{22} = 6$?

Solución: Resolviendo el determinante de K obtenemos:
 $|K| = (k_{11} \cdot k_{22} - k_{21} \cdot k_{12}) = (4 \cdot 6 - 2 \cdot 3) = 18$. Como $\text{mcd}(18, 27) = 3$, no puede usarse K como matriz de cifrado al carecer de inversa.

Con estos antecedentes, podremos entonces cifrar cualquier texto en claro mediante matrices usando una clave de dimensión $d \times d$ y agrupando el mensaje en bloques de tamaño d -gramas. Al igual que en otros sistemas, si es necesario se añaden al final del texto en claro caracteres de relleno para obtener el d -grama.

Ejemplo 1.48: Cifre el texto $M = \text{AMIGO CONDUCTOR, SI BEBES NO CONDUZCAS}$ mediante trigramas usando la matriz simbólica con clave PELIGROSO.

Solución: Primero vamos a comprobar que la matriz tiene inversa. El valor de $|K|$ en módulo 27 es igual a 4 como se indica:

$$|K| = [16(6 \cdot 15 - 19 \cdot 18) - 4(8 \cdot 15 - 15 \cdot 18) + 11(8 \cdot 19 - 15 \cdot 6)] \bmod 27$$

$|K| = [-9 + 6 + 7] \bmod 27 = 4$. Los valores de $|K|$ y el módulo n son primos entre sí, luego la clave es válida.

El mensaje M se divide en trigramas genéricos del tipo $M_i M_j M_k$ como se indica: AMI GOC OND UCT ORS IBE BES NOC OND UZC ASX que pasan a cifrarse con la matriz de clave simbólica PELIGROSO. Puesto que $(AMI) = (00 \ 12 \ 08)$ entonces el primer subcriptograma C_A será:

$$\begin{pmatrix} C_1 \\ C_2 \\ C_3 \end{pmatrix} = \begin{pmatrix} 16 & 4 & 11 \\ 8 & 6 & 18 \\ 15 & 19 & 15 \end{pmatrix} \begin{pmatrix} 0 \\ 12 \\ 8 \end{pmatrix} \bmod 27$$

Resolviendo esta matriz se obtiene:

$$C_1 = (16 \cdot 0 + 4 \cdot 12 + 11 \cdot 8) \bmod 27 = 136 \bmod 27 = 1$$

$$C_2 = (8 \cdot 0 + 6 \cdot 12 + 18 \cdot 8) \bmod 27 = 216 \bmod 27 = 0$$

$$C_3 = (15 \cdot 0 + 19 \cdot 12 + 15 \cdot 8) \bmod 27 = 348 \bmod 27 = 24$$

$$C_A = 01 \ 00 \ 27 = \text{BAX}$$

Los demás subcriptogramas se los dejo como ejercicio.

Cifrador de Hill digramico

Si un texto en claro $M = M_1 M_2 M_3 M_4 \dots M_N$ se cifra según el método de Hill en bloques de dos caracteres, para cada par de letras se tendrá:

$$C_1 = (k_{11} M_1 + k_{12} M_2) \bmod n \quad (1.55)$$

$$C_2 = (k_{21}M_1 + k_{22}M_2) \bmod n \quad (1.56)$$

$$\begin{pmatrix} C_1 \\ C_2 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} \begin{pmatrix} M_1 \\ M_2 \end{pmatrix} \bmod n \quad (1.57)$$

Ejemplo 1.49: Si la clave K es $k_{11} = 4$; $k_{12} = 2$; $k_{21} = 9$; $k_{22} = 2$, cifre el siguiente mensaje $M = \text{QUE TODA LA VIDA ES SUEÑO Y LOS SUEÑOS SUEÑOS SON.}$

Solución: $|K| = (k_{11}k_{22} - k_{21}k_{12}) \bmod 27 = (4 \cdot 2 - 2 \cdot 9) \bmod 27 = -10 \bmod 27$ que es igual a $17 \bmod 27$. El $\text{mcd}(17, 27) = 1$, luego existirá la matriz inversa.

Representamos el mensaje como se muestra y procedemos a cifrar el primer digrama con la matriz K ; los demás se los dejo como ejercicio.

$M = \text{QU ET OD AL AV ID AE SS UE ÑO YL OS SU EÑ OS SU EÑ OS SO NX.}$

Como $[C_1 C_2] = [K]x[M_1 M_2]$, para $[M_1 M_2] = \text{QU} = [17 \ 21]$ se tiene:

$C_1 = (4 \cdot 17 + 2 \cdot 21) \bmod 27 = 2 = C$ y $C_2 = (9 \cdot 17 + 2 \cdot 21) \bmod 27 = 6 = G$.

$C_1 C_2 = \text{CG}$. El criptograma completo que debe obtener es:

$C = \text{CGCV MGVV QQLX IIGT LIFU ÑEQL KXQK QLKX QKQL YMSD.}$

Para descifrar un criptograma de Hill, conocida la matriz clave, procedemos a calcular su inversa como ya se ha explicado utilizando la ecuación (1.48)

Ejemplo 1.50: Sea la matriz $[K]$: $k_{11}=2$, $k_{12}=10$, $k_{21}=17$, $k_{22}=5$. Se pide descifrar el siguiente criptograma $C = \text{NXXZ XSNX NEKE MJZT RVXD ÑZWB XZYW RJEV.}$

Solución: Cálculo de la matriz inversa K^{-1} :

$|K| = (2 \cdot 5 - 17 \cdot 10) \bmod 27 = 2$. El $\text{inv}(2, 27) = 14$.

$$K^{-1} = \frac{1}{|K|} \begin{pmatrix} k_{22} & -k_{12} \\ -k_{21} & k_{11} \end{pmatrix} \bmod 27 = \frac{1}{2} \begin{pmatrix} 5 & -10 \\ -17 & 2 \end{pmatrix} \bmod 27$$

$$K^{-1} = \begin{pmatrix} 14 \cdot 5 & -14 \cdot 10 \\ -14 \cdot 17 & 14 \cdot 2 \end{pmatrix} \bmod 27$$

$$K^{-1} = \begin{pmatrix} 16 & 22 \\ 5 & 1 \end{pmatrix} \bmod 27$$

Aplicando la matriz inversa K^{-1} al criptograma C se obtiene el siguiente mensaje $M = \text{HILL SE HIZO FAMOSO PERO NO MILLONARIO.}$

Compruebe este resultado y si es capaz de hacerlo sin el auxilio de una calculadora, mi enhorabuena por su agilidad mental.

1.7.4. Criptoanálisis del cifrado de Hill

Si la matriz clave, en donde descansa la seguridad del sistema, no puede contener todas las combinaciones posibles del Conjunto Completo de Restos del módulo de trabajo, puesto que eliminamos aquellos que nos entregan un determinante igual a cero o bien tienen factores comunes con n , ¿qué tan seguro es este criptosistema? Estudiaremos a continuación el número de las matrices válidas para determinar la entropía de la clave $H(K)$.

Supongamos, por simplicidad, que se trabaja en módulo 2, posteriormente lo

haremos en módulo 27. Si la matriz es de orden dos, es decir, tiene los elementos k_{11} , k_{12} , k_{21} y k_{22} , entonces el número posible de matrices será igual a 16 puesto que existirán 2^4 combinaciones del CCR módulo 2, es decir los valores 0 y 1. Estas matrices serán:

$$\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}; \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}; \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}; \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix} \\ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}; \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}; \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}; \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}; \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

No obstante, solamente existirán 6 matrices válidas, aquellas en las que el determinante es distinto de cero; en este caso en particular las matrices 7ª y 8ª de la primera fila y 2ª, 4ª, 6ª y 7ª de la segunda. Si, por ejemplo, utilizamos sólo tres números, los restos 0, 1 y 2 del módulo 27 como elementos de la matriz, se obtienen $3^4 = 81$ matrices de 2×2 . De estas matrices 33 serán no válidas por lo que sólo 48 matrices clave con los restos 0, 1 y 2 permiten cifrar digramas en dicho cuerpo. Si se atreve y tiene tiempo suficiente, compruébelo.

En el valor de módulo 27, tendremos $27^4 = 531.441$ matrices distintas de orden 2 cuyos elementos son el CCR(27), es decir $[0, 26]$. Si ahora descartamos aquellas matrices en las que el determinante es igual a cero o bien tienen factor común con el módulo 27, el número de matrices válidas se reduce a 314.928. Si deseamos aumentar la entropía de la clave, podríamos trabajar con un módulo primo, por ejemplo en módulo 37 con un alfabeto de letras mayúsculas más los dígitos del 0 al 9, de forma que en este caso prácticamente sólo se eliminen aquellas matrices cuyo determinante sea cero. Con un módulo igual a 37, el número de matrices 2×2 crece hasta 1.874.161 de las que más de 1.800.000 son claves válidas. No se asuste, no le pediré que compruebe esto.

En cuanto a la distancia de unicidad de este cifrador, que dependerá de la entropía de la clave o, lo que es lo mismo, del número de matrices válidas para cifrar, podemos aproximar de forma empírica para digramas y trigramas los siguientes valores cuando el módulo de trabajo n es un número primo:

$$\text{Digramas: } N = H(K)/D \approx \log_2[n^4 - n^3 - n^2 + n]/3,4 \quad \boxed{1.58}$$

$$\text{Trigramas: } N = H(K)/D \approx \log_2[n^9 - n^8 - n^7 + n^5 + n^4 - n^3]/3,4 \quad \boxed{1.59}$$

Por ejemplo, cifrando digramas en módulo 37 se obtendría una distancia de unicidad igual a 6,1 caracteres; aumentando el d-grama en una unidad, es decir cifrando trigramas, la distancia de unicidad crece hasta 13,8 que es más del doble de la anterior. Como el número de matrices tiende a $n(\exp d^2)$ para poligramas de longitud d , la distancia de unicidad aumentará significativamente. Por desgracia, esta característica no aumentará su nivel de seguridad como veremos más adelante.

El cifrador de Hill se muestra por tanto, al menos en una primera aproximación, bastante robusto ante un ataque puesto que, además, el algoritmo de cifrado destruye

las estadísticas del lenguaje y cuenta con una característica muy interesante en criptografía: el cifrado de los caracteres de un bloque dependerá también de los caracteres que forman el poligrama y de su posición relativa en él. Por ejemplo, si ciframos dos mensajes $M_A = OK$ y $M_B = OH$, en principio muy similares aunque signifiquen cosas distintas, el resultado no tendrá en cuenta para nada esta relación. Por ejemplo, si la matriz de cifra es $k_{11} = 2$, $k_{12} = 1$, $k_{21} = 1$ y $k_{22} = 3$, estos mensajes se cifrarán siguiendo la ecuación (1.56) como sigue:

$$\begin{aligned} M_A = OK & \Rightarrow C_1 = (15*2 + 10*1) \bmod 27 = 13 = N \\ & C_2 = (15*1 + 10*3) \bmod 27 = 18 = R \\ & \text{Luego } C = NR \end{aligned}$$

$$\begin{aligned} M_B = OH & \Rightarrow C_1 = (15*2 + 7*1) \bmod 27 = 10 = K \\ & C_2 = (15*1 + 7*3) \bmod 27 = 9 = J \\ & \text{Luego } C = KJ \end{aligned}$$

Por otra parte, si $M = KO$ el criptograma será $C = IB$. Luego, un simple cambio de posición de los caracteres en el texto en claro o modificaciones mínimas, producen una alteración total en el criptograma. Esto se debe a la ecuación genérica de cifra que, por ejemplo, para digramas es:

$$C_i = (k_{i1}*M_1 + k_{i2}*M_2) \bmod n \quad \boxed{1.60}$$

De la ecuación (1.60) se deduce que el carácter que ocupa la posición i ésima en el criptograma depende no sólo del carácter que ocupa la posición i ésima en el texto en claro, sino también del siguiente en la posición $i+1$ que conforma, en este caso, el digrama. Luego, mientras mayor sea el tamaño del poligrama utilizado, cada carácter dependerá de más caracteres del texto en claro y ahí radica en principio la fortaleza de este cifrado. En estas condiciones, no cabe plantearse un ataque por análisis de frecuencias. Por otra parte, el ataque por fuerza bruta puede ser extremadamente difícil si se elige un primo como módulo de trabajo y se cifran bloques de texto de un tamaño igual o mayor que 5.

Ejemplo 1.51: *Si el grupo de trabajo es el primo 37 y por tanto casi el 100% de las matrices de cifrado son válidas, ¿qué tamaño de poligrama debe usarse para que la entropía de la clave $H(K)$ del cifrado de Hill sea del orden de la del algoritmo DES calculada en el capítulo segundo e igual a 56?*

Solución: *Para $d=3$, la matriz clave K tendrá $3 \times 3 = 9$ elementos, por lo que el número de matrices puede estimarse en $37^9 \approx 1,3 \times 10^{14}$. Luego la entropía de la clave $H(K) = \log_2(1,3 \times 10^{14}) = 46,9$. Aumentando a tetragramas, el número de matrices es del orden de $37^{16} \approx 1,2 \times 10^{25}$ con lo que la entropía en este caso se eleva a $H(K) = \log_2(1,2 \times 10^{25}) = 83,3$. Luego, para un poligrama igual a 4 caracteres, este cifrado tendría una fortaleza similar al DES estándar en cuanto a la distancia de unicidad. ¿Y si ahora ciframos con bloques de texto en claro de 8 caracteres (64 bits) como lo hace el DES?*

Colmo es fácil apreciar, en estas condiciones un ataque por fuerza bruta es

impensable. La única posibilidad de ataque a este tipo de cifra es la elección de un texto en claro y buscar vectores unitarios en el mensaje o en el criptograma, y en el caso de no encontrarlos aplicar el método de Gauss-Jordan contando ahora sólo con un criptograma y su correspondiente texto en claro. En cualquier caso supondremos, además, que el criptoanalista conoce que el cifrado se trata de Hill, que conoce el tamaño del poligrama usado para la cifra y la correspondencia entre los caracteres del alfabeto en claro y su equivalente numérico. He aquí el verdadero *Talón de Aquiles* de este cifrador y la razón por la que, incluso alcanzando un valor de distancia de unicidad muy alto, no es seguro y por tanto ha caído en desuso por completo. A comienzos de este año 1999 ha vuelto a ponerse de moda como decíamos el uso de la matemática de matrices en la cifra, pero no se trata de este tipo de algoritmo.

• Ataque con elección del texto en claro o criptograma

¿Qué es eso de los vectores unitarios y el ataque aplicando el método de Gauss-Jordan? Ahora lo explicaremos. Vamos a suponer que el criptoanalista cuenta con el criptograma y los correspondientes textos en claro de varios mensajes. Luego, podrá elegir bloques específicos que le reporten mayor información. Para este tipo de cifra interesa encontrar los vectores unitarios de la dimensión en la que estamos trabajando. El resultado de cifrar estos poligramas serán los distintos valores de las columnas de la matriz clave como lo comprobaremos ahora mismo.

Un vector unitario de dimensión n es aquel que tiene todos sus elementos nulos excepto el elemento i ésimo que es la unidad. Por ejemplo, para una cifra con trigramas, $n = 3$, la matriz de Identidad I_3 tendrá los vectores unitarios μ_1 , μ_2 y μ_3 que se indican:

$$\text{Si } I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ entonces } \begin{aligned} \mu_1 &= [100] \\ \mu_2 &= [010] \\ \mu_3 &= [001] \end{aligned} \quad \boxed{1.61}$$

Supongamos entonces que realizamos la siguiente operación matricial:

$$K \times \mu_1 = \begin{pmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{pmatrix} \times \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} k_{11} \\ k_{21} \\ k_{31} \end{pmatrix} \quad \boxed{1.62}$$

Como se observa, resolviendo la ecuación (1.62) se obtiene la primera columna de la matriz clave. Si realizamos la misma operación con los vectores unitarios μ_2 y μ_3 encontramos la segunda y tercera columnas de dicha matriz. Por lo tanto, si ciframos con un vector unitario i , encontramos la columna i de la matriz clave que precisamente será el trigramas de texto cifrado que conoce el criptoanalista pues $K * M = C$.

Si el alfabeto de cifrado es el habitual, entonces los vectores unitarios para este caso trigramico serán:

$$\begin{aligned}\mu_1 &= [B A A] \\ \mu_2 &= [A B A] \\ \mu_3 &= [A A B]\end{aligned}$$

1.63

Ejemplo 1.52: Si se ha recibido el criptograma C y se conoce que pertenece al mensaje en claro M , se pide encontrar la matriz clave de cifrado digramico.

$M = EL BANDIDO FUE ABATIDO AL ATARDECER.$

$C = OYFCQ LSBEW FECEN ZSBUD BVSNO UXPCZ.$

Solución: Escribimos el mensaje y el criptograma en digramas:

$M = EL \underline{BA} ND ID OF UE \underline{AB} AT ID OA LA TA RD EC ER.$

$C = OY FC \underline{QL} SB EW FE CE NZ SB UD BV SN OU XP CZ.$

Encontramos los vectores unitarios $[A B]$ y $[B A]$ en los digramas segundo y séptimo. El cifrado correspondiente al vector $[B A]$ es FC , es decir los números 5 y 2, en tanto que el correspondiente al vector $[A B]$ es CE , es decir 2 y 4. Luego la matriz clave será:

$$K = \begin{pmatrix} 5 & 2 \\ 2 & 4 \end{pmatrix}$$

En el ejemplo anterior, para el primer digrama de texto en claro $BA = [1 \ 0]$ se tiene que $C_1 = k_{11}*1 + k_{12}*0 = k_{11} = 5 = F$ y $C_2 = k_{21}*1 + k_{22}*0 = k_{21} = 2 = C$, el digrama que aparece en el texto cifrado. Por lo tanto, si en el texto en claro el criptoanalista encuentra estas cadenas de vectores unitarios, será capaz de encontrar la matriz de cifrado. Esto será también válido para cifrados trigramicos. No obstante, para n mayor que tres, el método deja de ser válido pues existen pocas cadenas de ese tipo en castellano. El vector de longitud cuatro $AAAB$ podríamos encontrarlo en el texto "... así que ella estaba dispuesta a abanicarse por el calor que hacía..." pero esto es hilar muy fino porque nos faltaría encontrar otros vectores como $BAAA$, $ABAA$, y $AABA$. Para cinco seguro que no existen.

Si sólo encontramos un vector unitario, por ejemplo $[A B]$, sigue siendo posible descriptar la matriz clave digramica. De igual manera sucederá si para un cifrado trigramico encontramos dos vectores unitarios, por ejemplo $[A A B]$ y $[A B A]$. El restante vector se puede deducir aplicando la ecuación $C = K * M$ con las incógnitas del caso; como conocemos C y M , se pueden despejar las incógnitas de la columna de la matriz clave que falta.

Ejemplo 1.53: El mensaje $M = "Ese abanico estaba abajo"$ se cifra por trigramas según Hill y se obtiene el siguiente criptograma:

$M = ESE ABA NIC OES TAB AAB AJO$

$C = AEA DFI EJJ KTL QYÑ EGJ GAR$

Se pide encontrar la matriz clave.

Solución: En el texto en claro existen dos vectores unitarios, el vector $\mu_2 = [ABA]$ que se cifra como $[DFI]$ por lo que la segunda columna de la matriz clave será $[3 \ 5 \ 8]$ y luego $\mu_3 = [AAB]$ que se cifra como $[EGJ]$ y por tanto la tercera columna de la matriz será $[4 \ 6 \ 9]$. Luego se tendrá la siguiente matriz:

$$K = \begin{pmatrix} k_{11} & 3 & 4 \\ k_{21} & 5 & 6 \\ k_{31} & 8 & 9 \end{pmatrix}$$

Si tomamos, por ejemplo el primer trigramma se tiene que el texto en claro [ESE] = [4 19 4] se cifra como [AEA] = [0 4 0], luego se cumplirá que:

$$\begin{pmatrix} 0 \\ 4 \\ 0 \end{pmatrix} = \begin{pmatrix} k_{11} & 3 & 4 \\ k_{21} & 5 & 6 \\ k_{31} & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 4 \\ 19 \\ 4 \end{pmatrix}$$

Resolviendo:

$$0 = (k_{11} \cdot 4 + 3 \cdot 19 + 4 \cdot 4) \bmod 27 \Rightarrow k_{11} = 2$$

$$4 = (k_{21} \cdot 4 + 5 \cdot 19 + 6 \cdot 4) \bmod 27 \Rightarrow k_{21} = 5$$

$$0 = (k_{31} \cdot 4 + 8 \cdot 19 + 9 \cdot 4) \bmod 27 \Rightarrow k_{31} = 7$$

(véase la explicación a continuación del ejemplo)

Luego la matriz clave K será:

$$K = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

La elección del texto en claro [ESE] = [4 19 4] en el ejemplo anterior es la adecuada puesto que los valores de k_{11} , k_{21} y k_{31} se obtendrán al multiplicarse por el primer elemento, la letra E = 4, que tiene inverso en n. Puesto que $\text{inv}(4, 27) = 7$:

$$\begin{aligned} k_{11}: \quad 0 &= (k_{11} \cdot 4 + 73) \bmod 27 = (k_{11} \cdot 4 + 19) \bmod 27 \\ k_{11} &= (0 - 19) \cdot \text{inv}(4, 27) \bmod 27 = 8 \cdot 7 \bmod 27 = 56 \bmod 27 = 2 \quad \Rightarrow k_{11} = 2 \end{aligned}$$

$$\begin{aligned} k_{21}: \quad 4 &= (k_{21} \cdot 4 + 119) \bmod 27 = (k_{21} \cdot 4 + 11) \bmod 27 \\ k_{21} &= (4 - 11) \cdot \text{inv}(4, 27) \bmod 27 = 20 \cdot 7 \bmod 27 = 140 \bmod 27 = 5 \quad \Rightarrow k_{21} = 5 \end{aligned}$$

$$\begin{aligned} k_{31}: \quad 0 &= (k_{31} \cdot 4 + 188) \bmod 27 = (k_{31} \cdot 4 + 26) \bmod 27 \\ k_{31} &= (0 - 26) \cdot \text{inv}(4, 27) \bmod 27 = 1 \cdot 7 \bmod 27 = 7 \bmod 27 = 7 \quad \Rightarrow k_{31} = 7 \end{aligned}$$

A igual resultado llegaremos si tomamos en este ejemplo los pares mensaje criptograma NIC/EJL y TAB/QYÑ; no así si la elección es OES/KTL y AJO/GAR.

Ejemplo 1.54: Demuestre que se obtiene la misma primera columna de la matriz k_{11} , k_{21} y k_{31} del ejemplo anterior, eligiendo el par mensaje/criptograma NIC/EJL y que la elección de los pares EOS/KTL y AJO/GAR no es la adecuada.

Solución: a) El mensaje NIC tiene el equivalente numérico 13, 8, 2 y el criptograma EJL 4, 9, 11. Luego:

$$\begin{pmatrix} 4 \\ 9 \\ 11 \end{pmatrix} = \begin{pmatrix} k_{11} & 3 & 4 \\ k_{21} & 5 & 6 \\ k_{31} & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 13 \\ 8 \\ 2 \end{pmatrix}$$

$$4 = (k_{11} \cdot 13 + 3 \cdot 8 + 4 \cdot 2) \bmod 27$$

$$\begin{aligned}
k_{11} &= (4 - 32) * \text{inv}(13,27) \bmod 27 = 26 * 25 \bmod 27 = 650 \bmod 27 = 2 \\
9 &= (k_{21} * 13 + 5 * 8 + 6 * 2) \bmod 27 \\
k_{21} &= (9 - 52) * \text{inv}(13,27) \bmod 27 = 11 * 25 \bmod 27 = 275 \bmod 27 = 5 \\
11 &= (k_{31} * 13 + 8 * 8 + 9 * 2) \bmod 27 \\
k_{31} &= (11 - 82) * \text{inv}(13,27) \bmod 27 = 10 * 25 \bmod 27 = 250 \bmod 27 = 7
\end{aligned}$$

b) Para el par EOS/KTL se tiene:

$$\begin{pmatrix} 10 \\ 20 \\ 11 \end{pmatrix} = \begin{pmatrix} k_{11} & 3 & 4 \\ k_{21} & 5 & 6 \\ k_{31} & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 15 \\ 4 \\ 19 \end{pmatrix}$$

$$10 = (k_{11} * 15 + 3 * 4 + 4 * 19) \bmod 27$$

$$k_{11} = (10 - 88) * \text{inv}(15,27) \bmod 27$$

Como $\text{mcd}(15,27) = 3$, no existe inverso y no puede calcularse k_{11} .

c) Para el par AJO/GAR se tiene:

$$\begin{pmatrix} 6 \\ 0 \\ 18 \end{pmatrix} = \begin{pmatrix} k_{11} & 3 & 4 \\ k_{21} & 5 & 6 \\ k_{31} & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 0 \\ 9 \\ 15 \end{pmatrix}$$

$$6 = (k_{11} * 0 + 3 * 9 + 4 * 15) \bmod 27$$

$$k_{11} = (6 - 87) * \text{inv}(0,27) \bmod 27$$

Como no existe $\text{inv}(0,n)$ no puede calcularse k_{11} .

En el ejemplo anterior, se podría pensar abordar el punto b) mediante el método de prueba de valores de k_{11} en la ecuación $10 = (k_{11} * 15 + 88) \bmod 27$. No obstante, esto es un error como veremos a continuación. Evidentemente el valor $k_{11}=2$ (que es el valor verdadero) cumple con la ecuación anterior pero también se cumple dicha ecuación para los valores $k_{11} = 11$ y $k_{11} = 20$ lo cual no tiene sentido porque la solución debe ser única.

Ahora bien, si el texto en claro no cuenta con estos vectores unitarios, también podemos buscarlos en el criptograma. En este caso, el procedimiento nos lleva a recuperar la matriz inversa de la utilizada para cifrar.

Ejemplo 1.55: Se tiene el siguiente texto en claro y su criptograma que se sabe ha sido cifrado mediante Hill por digramas. Encuentre la matriz clave.

M = HILL SE PUEDE ATACAR Y ROMPER LA CIFRA BUSCANDO VECTORES
C = IBSD WJ QQLCL QBESVA B JOXH LI KN BHSR EOKSVCTM KTYBKAYFI

Solución: Ordenando por digramas tenemos:

M = HI LL SE PU ED EA TA CA RY RO MP ER LA CI FR AB US CA ND OV EC TO RE SX
C = IB SD WJ QQ LC LQ BE SV AB JO XH LI KN BA HS RE OK SV CT MK TY BK AY FI

En el criptograma están los dos vectores unitarios $[BA] = [1 \ 0]$ con su par en el texto en claro $[CI] = [2 \ 8]$ y el vector $[AB] = [0 \ 1]$ con su par de texto en claro $[RY] = [18 \ 25]$.

Luego, de acuerdo con las ecuaciones (1.62) y (1.63) se tiene que:

$$K^{-1} = \begin{pmatrix} 2 & 18 \\ 8 & 25 \end{pmatrix}$$

Como $K = (K^{-1})^{-1}$, se puede deducir que la matriz clave será entonces:

$$K = \begin{pmatrix} 23 & 18 \\ 11 & 4 \end{pmatrix}$$

Amigo lector(a); le dejo como ejercicio comprobar que con esta matriz de cifra K se obtiene el criptograma indicado.

Para el caso de cifra con digramas, buscamos algún digrama que contenga el valor cero o la letra A, bien en el mensaje en claro o bien en el criptograma; planteamos entonces un sistema de ecuaciones en donde la única condición a cumplir es que el elemento que acompañe a esa letra A tenga inverso en el cuerpo de cifra. De esta manera se obtiene una de las columnas de la matriz clave. Para encontrar la columna restante planteamos otra ecuación de cifra, en donde el elemento que multiplica a los k_{ij} deberán también tener inverso, como se mostrará en el siguiente ejemplo. ¿Y si no contamos con estos vectores unitarios? Aunque no lo crea, todavía podremos atacar al sistema si conocemos el texto en claro.

Ejemplo 1.56: *Se nos pide realizar un ataque al sistema de cifra de Hill digramico según el método explicado. El texto en claro y su criptograma son:*

M = HABIA VIDA EN MARTE

C = PIEBX PQYX YN FARIQ

Solución: *Agrupando texto en claro y criptograma por digramas:*

M = HA BI AV ID AE NM AR TE

C = PI EB XP QY XY NF AR IQ

Como no vemos vectores unitarios por ninguna parte, vamos a plantear la primera ecuación de cifra para el primer digrama HA en donde aparece la letra H = 7 en el texto en claro acompañado de la letra A = 0. Puesto que $\text{inv}(7,27) = 4$, entonces:

$$\begin{pmatrix} P \\ I \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} x \begin{pmatrix} H \\ A \end{pmatrix} \quad \begin{pmatrix} 16 \\ 8 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} x \begin{pmatrix} 7 \\ 0 \end{pmatrix}$$

$$16 = k_{11} * 7 \Rightarrow k_{11} = 16 * \text{inv}(7,27) \bmod 27 = 16 * 4 \bmod 27 = 10$$

$$8 = k_{21} * 7 \Rightarrow k_{21} = 8 * \text{inv}(7,27) \bmod 27 = 8 * 4 \bmod 27 = 5$$

*La ecuación del digrama (XP) = K * (AV), en donde V = 22 tiene como inverso 16, nos entrega la siguiente matriz:*

$$\begin{pmatrix} X \\ P \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} x \begin{pmatrix} A \\ V \end{pmatrix} \quad \begin{pmatrix} 24 \\ 16 \end{pmatrix} = \begin{pmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{pmatrix} x \begin{pmatrix} 0 \\ 22 \end{pmatrix}$$

$$24 = k_{12} * 22 \Rightarrow k_{12} = 24 * \text{inv}(22,27) \bmod 27 = 24 * 16 \bmod 27 = 6$$

$$16 = k_{22} * 22 \Rightarrow k_{22} = 16 * \text{inv}(22,27) \bmod 27 = 16 * 16 \bmod 27 = 13$$

Por lo tanto, la matriz de cifra del ejemplo es:

$$K = \begin{pmatrix} 10 & 6 \\ 5 & 13 \end{pmatrix}$$

Este método podría generalizarse para matrices de mayor rango, aunque como es lógico aumentará la dificultad de encontrar poligramas con todos los elementos excepto uno iguales a cero, por lo que este método resulta poco práctico. La generalización del ataque anterior mediante el planteamiento de un sistema de ecuaciones matriciales, se conoce como método de *Gauss-Jordan* y será tratado en el próximo apartado. En este otro escenario prácticamente no hay criptograma que se

resista a este ataque; no obstante, si se desconoce el texto en claro, debido al alto valor de la distancia de unicidad de este cifrador resulta absurdo intentar un ataque por fuerza bruta.

• Ataque con texto en claro conocido

Al no poder utilizar la técnica anterior porque no se encuentran los vectores unitarios en el texto en claro o en el criptograma, el criptoanalista siempre podrá atacar un cifrado de Hill si cuenta, por lo menos, con un criptograma y su texto en claro asociado. Siguiendo el método propuesto por *Alan Konheim* en "*Cryptography: A Primer*"⁶ el procedimiento consiste en disponer las correspondencias entre el texto en claro y el texto cifrado en forma de matriz y utilizar el método de Gauss-Jordan que consiste, básicamente, en aplicar operaciones elementales a la matriz hasta conseguir (si se puede) diagonalizar la parte izquierda, de forma que la diagonal principal sea la unidad. Esto quiere decir que en la mitad izquierda tendremos los vectores unitarios que definimos en el apartado anterior, por lo que la otra mitad derecha tendrá una relación directa con la matriz clave buscada.

Si la matriz $2n$ -grámica la definimos como $[(\text{TextoEnClaro}) \mid (\text{TextoCifrado})]$ la parte derecha, una vez diagonalizada la izquierda, será la traspuesta de la matriz clave de cifrado K . Le dejo aquí un nuevo ejercicio: definir ahora la matriz $2n$ -grámica como $[(\text{TextoCifrado}) \mid (\text{TextoEnClaro})]$ y una vez diagonalizada la parte izquierda comprobar qué tipo de matriz se obtiene. Por ejemplo, suponga que tenemos el siguiente texto en claro asociado con el criptograma de Hill trigrámico que se indica:

$$\begin{array}{l}
 M = \text{ENU NLU GAR DEL AMA NCH ADE CUY ONO MBR} \dots \\
 C = \text{WVX IDQ DDO ITQ JGO GJI YMG FVC UÑT RLL} \dots
 \end{array}$$

$$\begin{array}{l}
 \text{Matriz Trigrámica} \\
 \text{Texto en Claro} \\
 \text{Texto Cifrado}
 \end{array}
 \begin{pmatrix}
 E & N & U & & W & V & X \\
 N & L & U & & I & D & Q \\
 G & A & R & & D & D & O \\
 D & E & L & & I & T & Q \\
 A & M & A & & J & G & O \\
 N & C & H & & G & J & I \\
 A & D & E & & Y & M & G \\
 C & U & Y & & F & V & C \\
 O & N & O & & U & \tilde{N} & T \\
 M & B & R & & R & L & L
 \end{pmatrix}
 =
 \begin{pmatrix}
 4 & 13 & 21 & & 23 & 22 & 24 \\
 13 & 11 & 21 & & 8 & 3 & 17 \\
 6 & 0 & 18 & & 3 & 3 & 15 \\
 3 & 4 & 11 & & 8 & 20 & 17 \\
 0 & 12 & 0 & & 9 & 6 & 15 \\
 13 & 2 & 7 & & 6 & 9 & 8 \\
 0 & 3 & 4 & & 25 & 12 & 6 \\
 2 & 21 & 25 & & 5 & 22 & 2 \\
 15 & 13 & 15 & & 21 & 14 & 20 \\
 12 & 1 & 18 & & 18 & 11 & 11
 \end{pmatrix}$$

Figura 1.43. Matriz $2n$ -grámica de Gauss-Jordan del ejemplo.

Como se observa, no aparecen vectores trigrámicos unitarios ni en el texto en

⁶ Konheim, Alan G., "*Cryptography: A Primer*", John Wiley & Sons, 1981, pp. 116-120.

claro ni en el texto cifrado por lo que intentaremos el ataque por Gauss-Jordan. Escribimos entonces los trigramas del texto en claro a la izquierda y los del criptograma a la derecha en una matriz 2 n-grámica, con los correspondientes equivalentes numéricos como se indica en la Figura 1.43.

El primer paso será conseguir que toda la primera columna sea 0 excepto el elemento a_{11} ; para ello multiplicamos la fila primera por 7 ya que $\text{inv}(4, 27) = 7$ con lo que se tiene $(4*7 \ 13*7 \ 21*7 \ 23*7 \ 22*7 \ 24*7) \bmod 27 = (1 \ 10 \ 12 \ 26 \ 19 \ 6)$.

Si el primer elemento de la fila (en este caso $E = 4$) tuviera algún factor común con el módulo 27, el método sigue siendo válido porque en ese caso se moverían filas enteras y alguna habrá cuyo primer elemento sea primo relativo con el módulo. La matriz no cambiará; es más, a nivel matemático da lo mismo donde estén localizadas las filas, no así para los nosotros los humanos que interpretamos lo que allí está escrito. Hecho esto se realizan las siguientes operaciones básicas módulo 27:

- a) $2^{\text{a}} \text{ fila} = 2^{\text{a}} \text{ fila} - 13 * 1^{\text{a}} \text{ fila}$
- b) $3^{\text{a}} \text{ fila} = 3^{\text{a}} \text{ fila} - 6 * 1^{\text{a}} \text{ fila}$
- c) $4^{\text{a}} \text{ fila} = 4^{\text{a}} \text{ fila} - 3 * 1^{\text{a}} \text{ fila}$
- d) $6^{\text{a}} \text{ fila} = 6^{\text{a}} \text{ fila} - 13 * 1^{\text{a}} \text{ fila}$
- e) $8^{\text{a}} \text{ fila} = 8^{\text{a}} \text{ fila} - 2 * 1^{\text{a}} \text{ fila}$
- f) $9^{\text{a}} \text{ fila} = 9^{\text{a}} \text{ fila} - 15 * 1^{\text{a}} \text{ fila}$
- g) $10^{\text{a}} \text{ fila} = 10^{\text{a}} \text{ fila} - 12 * 1^{\text{a}} \text{ fila}$

Se obtiene entonces la siguiente matriz:

| | |
|--------------------------|---|
| | $\begin{pmatrix} 1 & 10 & 12 & 26 & 19 & 6 \\ 0 & 16 & 0 & 21 & 26 & 20 \\ 0 & 21 & 0 & 9 & 24 & 6 \\ 0 & 1 & 2 & 11 & 17 & 26 \\ 0 & 12 & 0 & 9 & 6 & 15 \\ 0 & 7 & 13 & 19 & 5 & 11 \\ 0 & 3 & 4 & 25 & 12 & 6 \\ 0 & 1 & 1 & 7 & 11 & 17 \\ 0 & 25 & 24 & 9 & 26 & 11 \\ 0 & 16 & 9 & 3 & 26 & 20 \end{pmatrix}$ |
| <i>Matriz Trigrámica</i> | |
| <i>Texto en Claro</i> | |
| <i>Texto Cifrado</i> | |

Procedemos de igual manera con las columnas segunda y tercera. Observe que en el cálculo de la columna tercera hemos tenido que mover filas porque aparece el valor 0 en el tercer elemento. Como ejercicio, compruebe qué movimientos se han hecho.

Al final de todo el proceso obtenemos la matriz 2n-grámica que se indica en donde se observan los vectores unitarios en la matriz de la izquierda, correspondiente

al texto en claro.

$$\begin{array}{l}
 \text{Matriz Trigrámica} \\
 \text{Texto en Claro} \\
 \text{Texto Cifrado} \\
 \text{diagonalizada} \\
 \text{agrupando los} \\
 \text{vectores unitarios}
 \end{array}
 \begin{pmatrix}
 1 & 0 & 0 & 2 & 5 & 7 \\
 0 & 1 & 0 & 3 & 5 & 8 \\
 0 & 0 & 1 & 4 & 6 & 9 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0 \\
 0 & 0 & 0 & 0 & 0 & 0
 \end{pmatrix}$$

Como la mitad izquierda correspondía al texto en claro, la parte derecha de la matriz con vectores unitarios será la traspuesta de la clave. Esto es, 100 es el primer vector unitario y entonces entrega la primera columna de la matriz de clave; y de igual manera sucede con los vectores segundo 010 y tercero 001 según la ecuación (1.63). Luego, la clave será:

$$K = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix}$$

Comprobemos que esta es la matriz verdadera, cifrando el primer trigrama del mensaje $M = [ENU] = [4 \ 13 \ 21]$ que debe darnos el trigrama $C = [WVX] = [23 \ 22 \ 24]$.

$$C = K \times M = \begin{pmatrix} 2 & 3 & 4 \\ 5 & 5 & 6 \\ 7 & 8 & 9 \end{pmatrix} \times \begin{pmatrix} 4 \\ 13 \\ 21 \end{pmatrix}$$

$$C_1 = [2*4 + 3*13 + 4*21] \bmod 27 = 23 = W$$

$$C_2 = [5*4 + 5*13 + 6*21] \bmod 27 = 22 = V$$

$$C_3 = [7*4 + 8*13 + 9*21] \bmod 27 = 24 = X$$

Le dejo como ejercicio comprobar la cifra completa del mensaje de este ejemplo. Atrévase a deducir la clave a partir de $[(\text{TextoCifrado}) \mid (\text{TextoEnClaro})]$.

1.8 CIFRADORES POR TRANSPOSICIÓN

El segundo método clásico utilizado para cifrar mensajes es la *transposición* o *permutación* de caracteres. Esto consiste en reordenar los caracteres del texto en claro

como si de una baraja de cartas se tratase. El resultado de tal acción es la de difuminar la información del texto en claro y provocar, por tanto, la difusión propuesta por Shannon para la protección de la misma. Precisamente este método era el utilizado por los lacedemonios en el sistema de cifra de la escítala que vimos en el apartado 1.1.1, cifrador que podríamos clasificar en la categoría de transposición por grupos.

Debe tenerse presente que al reordenar el texto, en el criptograma aparecerán exactamente los mismos caracteres que en el texto en claro y que, por tanto, no evitamos en este caso que un intruso detecte fácilmente que nuestro criptosistema es de transposición mediante la simple acción de contabilizar los caracteres del texto cifrado y comparar las frecuencias relativas con las del lenguaje. Esto es, si en el alfabeto de 27 letras la letra *E* aparece cerca del 13%, la letra *A* cerca del 10%, etc., no cabe duda que el cifrado ha sido realizado por permutaciones. No obstante, sí se destruyen los digramas, trigramas y, en general poligramas, al separar los caracteres en el texto cifrado.

Ahora bien, aunque se detecte una distribución de caracteres en el criptograma muy parecida a la característica del lenguaje, sólo nos indica eso, que es muy posible que se haya cifrado por transposiciones, pero de nada nos servirá la técnica utilizada en cifradores por sustitución para intentar un criptoanálisis. En este caso, el ataque deberá plantearse con el uso de una técnica denominada *anagramación* y que consiste en la comparación de bloques de caracteres del criptograma con el objeto de buscar la formación de los poligramas destruidos por el cifrado.

1.8.1. Transposición por grupos

En este tipo de cifra, los caracteres del texto en claro se reordenan por medio de una permutación $\Pi_x(y)$ en donde x indica la acción ejercida sobre el conjunto de caracteres del mensaje M e y es la posición ordenada de los caracteres según la acción x . Luego, si $a_1, a_2, a_3, \dots, a_k$ son letras del texto en claro, y Π es una permutación de $1, 2, 3, \dots, k$ números, entonces cada carácter C_i del criptograma será el resultado de aplicar dicha permutación sobre ese conjunto de k caracteres. Por ejemplo, sea el conjunto de números $[1, 10] = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10$ y sean x_1 y x_2 dos acciones de permutación tales que x_1 es la acción de ordenar cada grupo de diez caracteres del mensaje de forma que primero envía los caracteres impares y luego los pares y x_2 es la función de ordenar los caracteres del mensaje desde la posición mayor a la menor, entonces se tiene que:

$$\begin{aligned}\Pi_1 &= 1, 3, 5, 7, 9, 2, 4, 6, 8, 10. \\ \Pi_2 &= 10, 9, 8, 7, 6, 5, 4, 3, 2, 1.\end{aligned}$$

Entonces de los grupos indicados, tenemos por ejemplo que:

$$\Pi_1(4) = 7; \Pi_1(9) = 8; \Pi_2(1) = 10; \Pi_2(3) = 8$$

La transposición por grupos será periódica, de período p , tras el cual la permutación aplicada al texto en claro se repite. Esto es, si el mensaje que se desea cifrar $M = m_1m_2m_3 \dots m_{10}m_{11}m_{12}$ y la permutación aplicada con período 4 es $\Pi_M = 4132$,

entonces el criptograma generado será $C = m_4m_1m_3m_2m_8m_5m_7m_6m_{12}m_9m_{11}m_{10}$.

Ejemplo 1.57: *Utilizando la permutación $\Pi_M = 24531$ cifre el siguiente mensaje:*
 $M = \text{MANOS ARRIBA, ESTO ES UN ATRACO.}$

Solución: *Aplicando $\Pi_M = 24531$ al texto en claro obtenemos:*
 $M = \text{MANOS ARRI B AESTO ESUNA TRACO}$
 $C = \text{AOSNM RIBRA ETOSA SNAUE RCOAT}$

Si el período es pequeño, como en el ejemplo anterior, el criptograma podría atacarse fácilmente mediante técnicas de anagramación que veremos más adelante. Una solución a este problema podría consistir en hacer crecer el período de la transposición. En esta línea podríamos llegar a la situación límite en que el período es tan largo como el propio mensaje, dando lugar a los denominados cifradores de transposición por series.

1.8.2. Transposición por series

Esta técnica consiste en ordenar el mensaje como una cadena de submensajes, de forma que el mensaje original se transmite como $M' = M_{S_1}M_{S_2}M_{S_3}\dots$, en donde cada una de las cadenas sigue una función o serie; por ejemplo, M_{S_1} puede corresponder a los múltiplos de 3, M_{S_2} los números primos, M_{S_3} los números pares, etc.

Supongamos entonces un mensaje M con un total de 25 caracteres. Si se utilizan las 3 series M_{S_1} , M_{S_2} y M_{S_3} que se indican en ese mismo orden:

M_{S_1} : Relación de números primos
 M_{S_2} : Relación de números pares
 M_{S_3} : Relación de números impares

entonces la cifra se realizará como sigue:

$M = m_1m_2m_3m_4m_5m_6m_7m_8m_9m_{10}m_{11}m_{12}m_{13}m_{14}m_{15}m_{16}m_{17}m_{18}m_{19}m_{20}m_{21}m_{22}m_{23}m_{24}m_{25}$
 $M' = M_{S_1}M_{S_2}M_{S_3}$
 $M_{S_1} = 1, 2, 3, 5, 7, 11, 13, 17, 19, 23$
 $M_{S_2} = 4, 6, 8, 10, 12, 14, 16, 18, 20, 22, 24$
 $M_{S_3} = 9, 15, 21, 25$
 $C = m_1m_2m_3m_5m_7m_{11}m_{13}m_{17}m_{19}m_{23}m_4m_6m_8m_{10}m_{12}m_{14}m_{16}m_{18}m_{20}m_{22}m_{24}m_9m_{15}m_{21}m_{25}$

Al no tener período, este algoritmo de cifrado posee una mayor fortaleza pues dificulta el criptoanálisis, residiendo su seguridad en el secreto de las series utilizadas. No obstante, es necesario recorrer el texto en claro completo por lo que el método es muy lento.

Ejemplo 1.58: *Utilizando las series $M_{S_1}M_{S_2}M_{S_3}$ vistas anteriormente y en ese orden, cifre el mensaje $M = \text{ERRAR ES HUMANO, PERDONAR DIVINO.}$*

Solución: *El mensaje tiene 27 caracteres. Si se transmite la secuencia M_{S_1} , luego M_{S_2} y finalmente M_{S_3} , tenemos los siguientes bloques:*
 $M_{S_1} = 1,2,3,5,7,11,13,17,19,23$
 $M_{S_2} = 4,6,8,10,12,14,16,18,20,22,24,26$

$M_{S3} = 9,15,21,25,27$

El mensaje ordenado según las posiciones de los caracteres es:

1234567890 1234567890 1234567

ERRARESHUM ANOPERDONA RDIVINO

Permutando los caracteres según la serie $M_{S3}M_{S2}M_{S1}$ se obtiene:

$C = ERRRS AODNI AEHMN PROAD VNUER IO$

1.8.3. Transposición por columnas

• Cifrador de transposición por columnas simple

En este tipo de cifrados, se reordenan los caracteres del texto en claro en N_c columnas de forma que el mensaje así escrito se transmite luego por columnas, obteniéndose de esta manera el criptograma. El efecto, al igual que en los demás cifradores por permutación, es desplazar las letras de las posiciones adyacentes. Por ejemplo, si $N_c = 6$, la columna de cifrados podría quedar como se indica:

| Columna de cifrados | | | | | |
|---------------------|----------|----------|----------|----------|----------|
| C_1 | C_2 | C_3 | C_4 | C_5 | C_6 |
| C_7 | C_8 | C_9 | C_{10} | C_{11} | C_{12} |
| C_{13} | C_{14} | C_{15} | C_{16} | C_{17} | C_{18} |
| C_{19} | C_{20} | C_{21} | C_{22} | ... | ... |

Luego, el criptograma se obtiene leyendo de arriba hacia abajo en las columnas, es decir:

$$C = C_1C_7C_{13}C_{19} \dots C_2C_8C_{14}C_{20} \dots C_3C_9C_{15}C_{21} \dots \dots C_6C_{12}C_{18} \dots \quad \boxed{1.64}$$

Para proceder a la función de cifra, primero se busca una cuadrícula en función del tamaño del bloque del mensaje. Si en la búsqueda de dicha cuadrícula quedan espacios en blanco, éstos se rellenan con algún carácter nulo previamente determinado y que conocen el transmisor y el receptor del mensaje, por ejemplo la letra X.

Ejemplo 1.59: *Cifre el siguiente texto mediante transposición por columnas con $N_c = 6$. Se usará como carácter de relleno la letra X.*

$M = NUNCA ES TARDE CUANDO LA DICHA ES BUENA.$

Solución: *Escribimos el texto en columnas como se indica:*

| | | | | | |
|---|---|---|---|---|---|
| N | U | N | C | A | E |
| S | T | A | R | D | E |
| C | U | A | N | D | O |
| L | A | D | I | C | H |
| A | E | S | B | U | E |
| N | A | X | X | X | X |

Luego leyendo por columnas el criptograma resultante será:

$C = NSCLA NUTUA EANA DSXCR NIBXA DDCUX EEOHE X.$

Para descifrar un criptograma por columnas, el receptor primero calculará el número de filas N_f a partir de la longitud del texto cifrado L_c y el número de columnas

N_C , clave secreta que sólo él conoce.

$$N_F = L_C / N_C \quad 1.65$$

Hecho esto, escribe el texto cifrado de forma vertical en tantas filas como indique el valor de N_F y procede a leerlo por filas.

Ejemplo 1.60: Se ha recibido el criptograma $C = PLXIU IEESN GTSOO OEX$ y se sabe que ha sido cifrado en 6 columnas. Descifrelo.

Solución: Como el criptograma tiene $L_C = 18$ caracteres y se ha cifrado con $N_C = 6$, entonces $N_F = 18/6 = 3$. Escribimos las seis columnas con longitud de tres caracteres, es decir PLX, IUI, EES, NGT, SOO y OEX.

| | | | | | |
|---|---|---|---|---|---|
| P | I | E | N | S | O |
| L | U | E | G | O | E |
| X | I | S | T | O | X |

Luego leyendo por filas y descartando los caracteres de relleno al final de la matriz, se obtiene el profundo mensaje $M = PIENSO, LUEGO EXISTO$.

En general, para un cifrado por transposición a través de una matriz de dimensiones $j \times k$ (j filas y k columnas), existe una relación funcional entre el texto en claro y el criptograma. En el caso de la cifra por columnas, el carácter del texto en claro de la posición i ésima se desplaza a la posición $E_t(i)$ debido a la acción de permutación, en donde:

$$E_t(i) = j * [(i-1) \bmod k] + \text{trunc}[(i-1)/k] + 1 \quad 1.66$$

Donde $\text{trunc}(f(x))$ es la parte entera de la operación efectuada sobre $f(x)$. Luego, el carácter M_i del texto en claro se representará en el criptograma como:

$$M_i = C_{j * [(i-1) \bmod k] + \text{trunc}[(i-1)/k] + 1} \quad 1.67$$

Aplicaremos la ecuación (1.66) al texto anterior $M = PIENSO, LUEGO EXISTO$. La posición relativa de los caracteres en el mensaje y en el criptograma será:

| | | | | | | | | | | | | | | | | | | |
|----|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|
| i: | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 |
| M: | P | I | E | N | S | O | L | U | E | G | O | E | X | I | S | T | O | |
| C: | P | L | X | I | U | I | E | E | S | N | G | T | S | O | O | O | E | X |

Por ejemplo, se observa que las letras del texto en claro N ($i=4$), L ($i=7$), U ($i=8$), G ($i=10$), T ($i=16$) y el digrama EX ($i,k=12,13$) del texto en claro se desplazan, a las posiciones 10, 2, 5, 11, 12, 17 y 3 respectivamente. Como el receptor conoce el número de columnas N_C empleadas en el cifrado y la longitud L_C del criptograma, deduce que el número de filas $N_F = j$ es igual a $L_C / N_C = 18/6 = 3$. Comprobaremos a continuación estos valores en particular aplicando la ecuación (1.66).

| | |
|----|---|
| N: | $E_t(4) = 3 * [(4-1) \bmod 6] + \text{trunc}[(4-1)/6] + 1 = 10 \Rightarrow C_{10}$ |
| L: | $E_t(7) = 3 * [(7-1) \bmod 6] + \text{trunc}[(7-1)/6] + 1 = 02 \Rightarrow C_2$ |
| U: | $E_t(8) = 3 * [(8-1) \bmod 6] + \text{trunc}[(8-1)/6] + 1 = 05 \Rightarrow C_5$ |
| G: | $E_t(10) = 3 * [(10-1) \bmod 6] + \text{trunc}[(10-1)/6] + 1 = 11 \Rightarrow C_{11}$ |

$$\begin{aligned} T: & E_t(16) = 3*[(16-1)\text{mod}6] + \text{trunc}[(16-1)/6] + 1 = 12 \Rightarrow C_{12} \\ E: & E_t(12) = 3*[(12-1)\text{mod}6] + \text{trunc}[(12-1)/6] + 1 = 17 \Rightarrow C_{17} \\ X: & E_t(13) = 3*[(13-1)\text{mod}6] + \text{trunc}[(13-1)/6] + 1 = 03 \Rightarrow C_3 \end{aligned}$$

Ejemplo 1.61: *Aplicando la ecuación (1.66) cifre en columnas usando una clave $N_C = 3$ el mensaje $M = LA VIDA ES UNA TÓMBOLA$.*

Solución: *Como el texto en claro tiene 18 caracteres y el número de columnas $k = 3$, entonces el número de filas $j = 6$. Las posiciones de los caracteres en el criptograma serán:*

$$\begin{aligned} E_t(1) &= 6*[(1-1)\text{mod}3] + \text{trunc}[(1-1)/3] + 1 = 1 \Rightarrow C_1 = M_1 = L \\ E_t(2) &= 6*[(2-1)\text{mod}3] + \text{trunc}[(2-1)/3] + 1 = 7 \Rightarrow C_7 = M_2 = A \\ E_t(3) &= 6*[(3-1)\text{mod}3] + \text{trunc}[(3-1)/3] + 1 = 13 \Rightarrow C_{13} = M_3 = V \\ E_t(4) &= 6*[(4-1)\text{mod}3] + \text{trunc}[(4-1)/3] + 1 = 2 \Rightarrow C_2 = M_4 = I \\ E_t(5) &= 6*[(5-1)\text{mod}3] + \text{trunc}[(5-1)/3] + 1 = 8 \Rightarrow C_8 = M_5 = D \\ E_t(6) &= 6*[(6-1)\text{mod}3] + \text{trunc}[(6-1)/3] + 1 = 14 \Rightarrow C_{14} = M_6 = A \end{aligned}$$

Si continuamos la cifra se obtiene finalmente:

$C = LIENO OADSA MLVAU TBA$.

De igual manera, utilizando ahora la ecuación (1.67) podemos encontrar la posición que ocupaban los caracteres del texto en claro a partir del criptograma. Por ejemplo los seis primeros caracteres descifrados del mensaje PIENSO LUEGO EXISTO serán:

$$\begin{aligned} M_1 &= C_{3*[(1-1)\text{mod}6] + \text{trunc}[(1-1)/6] + 1} = C_1 \Rightarrow M_1 = P \\ M_2 &= C_{3*[(2-1)\text{mod}6] + \text{trunc}[(2-1)/6] + 1} = C_4 \Rightarrow M_2 = I \\ M_3 &= C_{3*[(3-1)\text{mod}6] + \text{trunc}[(3-1)/6] + 1} = C_7 \Rightarrow M_3 = E \\ M_4 &= C_{3*[(4-1)\text{mod}6] + \text{trunc}[(4-1)/6] + 1} = C_{10} \Rightarrow M_4 = N \\ M_5 &= C_{3*[(5-1)\text{mod}6] + \text{trunc}[(5-1)/6] + 1} = C_{13} \Rightarrow M_5 = S \\ M_6 &= C_{3*[(6-1)\text{mod}6] + \text{trunc}[(6-1)/6] + 1} = C_{16} \Rightarrow M_6 = O \end{aligned}$$

Ejemplo 1.62: *Descifre a través de la ecuación (1.67) el siguiente criptograma cifrado con $N_C = 4$ columnas. $C = CNEAM SAANY IXMNO CNXIT HAOX$.*

Solución: *Como $N_C = 4$ y el criptograma tiene $L_C = 24$ caracteres, obtenemos $j = 6$.*

$$\begin{aligned} M_1 &= C_{6*[(1-1)\text{mod}4] + \text{trunc}[(1-1)/4] + 1} = C_1 \Rightarrow M_1 = C \\ M_2 &= C_{6*[(2-1)\text{mod}4] + \text{trunc}[(2-1)/4] + 1} = C_7 \Rightarrow M_2 = A \\ M_3 &= C_{6*[(3-1)\text{mod}4] + \text{trunc}[(3-1)/4] + 1} = C_{13} \Rightarrow M_3 = M \\ M_4 &= C_{6*[(4-1)\text{mod}4] + \text{trunc}[(4-1)/4] + 1} = C_{19} \Rightarrow M_4 = I \\ M_5 &= C_{6*[(5-1)\text{mod}4] + \text{trunc}[(5-1)/4] + 1} = C_2 \Rightarrow M_5 = N \end{aligned}$$

Siguiendo con el mismo procedimiento y (hágalo Ud. mismo) obtenemos el siguiente texto en claro $M = CAMINANTE NO HAY CAMINOS$.

Como veremos en el próximo apartado, por mucho que con esta operación se destruyan poligramas, mediante una técnica denominada anagramación seremos capaces de atacar el criptograma. Esto es posible ya que en el cifrado anterior pueden persistir adyacencias de series cortas de letras, por ejemplo digramas característicos, desplazados una distancia constante. Ante ello, existen dos soluciones a este problema: aplicar una doble transposición o bien hacer uso de una clave para permutar las columnas antes de escribir el criptograma.

• Cifrador de transposición por columnas simple con clave

Para evitar o hacer más difícil el ataque por anagramación, podemos utilizar una clave con el objeto de cambiar la posición relativa de las columnas de la cuadrícula. Esta clave puede ser cualquier combinación de números desde 1 hasta N_C , no obstante podemos asociar una palabra de longitud N_C con todos los caracteres distintos a dicha combinación de números. Por ejemplo si se trabaja con 7 columnas y se desea una permutación de éstas del tipo 2547136, una posible palabra clave sería la palabra PERMISO pues, ordenando los caracteres de dicha clave alfabéticamente, se obtiene precisamente esa permutación: EIMOPRS.

Ejemplo 1.63: *Cifre por columnas con la clave RELOJ el siguiente mensaje.*

M = EL PATIO DE MI CASA ES PARTICULAR, CUANDO LLUEVE SE MOJA COMO LOS DEMÁS.

Solución: *Escribiendo el mensaje en 5 columnas y luego permutando éstas según la clave RELOJ, tenemos:*

| R | E | L | O | J | E | J | L | O | R |
|---|---|---|---|---|---|---|---|---|---|
| E | L | P | A | T | L | T | P | A | E |
| I | O | D | E | M | O | M | D | E | I |
| I | C | A | S | A | C | A | A | S | I |
| E | S | P | A | R | S | R | P | A | E |
| T | I | C | U | L | I | L | C | U | T |
| A | R | C | U | A | R | A | C | U | A |
| N | D | O | L | L | D | L | O | L | N |
| U | E | V | E | S | E | S | V | E | U |
| E | M | O | J | A | M | A | O | J | E |
| C | O | M | O | L | O | L | M | O | C |
| O | S | D | E | M | S | M | D | E | O |
| A | S | X | X | X | S | X | X | X | A |

Escribiendo las columnas resultantes, se tiene:

C = LOCSI RDEMO SSTMA RLALS ALMXP DAPCC OVOMD XAESA UULEJ OEXEI IETAN UECOA.

Si se desea provocar una mayor confusión y difusión en el criptograma, en otras palabras *rizar el rizo*, podríamos incluir por ejemplo un par de líneas más en la matriz después del fin del mensaje, utilizando las mismas letras del texto. De esta manera, si tomamos como caracteres de *relleno* los de las columnas 1ª, 3ª y 5ª del mensaje escrito en columnas antes de aplicar la clave, las últimas cuatro filas de la primera matriz serán ahora:

| | | | | |
|---|---|---|---|---|
| • | • | • | • | • |
| O | S | D | E | M |
| A | S | E | P | T |
| I | D | M | I | A |
| A | E | P | R | T |

El criptograma, que se lo dejo como ejercicio, será:

C = LOCSI RDEMO SSDET MARLA LSALM TATPD APCCO VOMDE MPAES AUULE JOEPI REIIE TANUE COAIA.

• Cifrador de doble transposición por columnas

Para destruir la adyacencia de series cortas de caracteres que pueden aparecer en una única transposición, también podemos utilizar una segunda permutación. Con ello, el criptograma final se obtiene tras aplicar las siguientes transformaciones:

$$C' = E_1(M) \quad 1.68$$

$$C = E_2[E_1(M)] \quad 1.69$$

Esto es, se escribe el texto del mensaje en claro M en columnas (operación E_1) en una matriz de dimensiones $j' \times k'$ y luego se reordena dicha matriz en otra de dimensión $j \times k$. El efecto de esta doble transposición será separar aún más los caracteres adyacentes y destruir, por tanto, los digramas. Compruebe Ud. mismo que al aplicar una doble permutación se produce una mayor dispersión de los caracteres del texto en claro en el criptograma final. En este caso ya no nos servirá el método de anagramación que veremos más adelante como herramienta de ataque a la cifra.

1.8.4. Transposición por filas

De forma similar al sistema de cifra por columnas, en esta operación de cifra se escribe el mensaje M en forma vertical, por ejemplo de arriba hacia abajo, con un cierto número de filas N_F que será la clave y luego se lee el criptograma en forma horizontal tal como se indica en el siguiente ejemplo.

Ejemplo 1.64: *Cifre por transposición de filas con clave $N_F = 3$ el siguiente mensaje:
 $M = EL PRISIONERO SE ENTREGARÁ EN EL LUGAR YA INDICADO.$*

Solución: Escribiendo el texto verticalmente en tres niveles, se tiene:

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| E | R | I | E | S | N | E | R | N | L | A | A | D | A |
| L | I | O | R | E | T | G | A | E | U | R | I | I | D |
| P | S | N | O | E | R | A | E | L | G | Y | N | C | O |

El criptograma se obtiene escribiendo las tres filas resultantes:

$C = ERIES NERNL AADAL IORET GAEUR IIDPS NOERA ELGYN CO.$

Evidentemente, las operaciones de cifra y descifrado serán análogas a las vistas en los sistemas de cifra por columnas. Esto es, conocido el número de elementos del criptograma L_C y la clave N_F , calculamos ahora el número de columnas N_C como L_C/N_F y luego se escribe el criptograma de forma horizontal en tantas columnas como sea el valor de N_C encontrado. Leyendo el resultado por columnas, en forma vertical de arriba hacia abajo, se obtiene el texto en claro.

Ejemplo 1.65: *Descifre el siguiente criptograma de cifra por filas y clave $N_F = 3$.
 $C = MAPDD ITOOE RURNX.$*

Solución: *Como la longitud del criptograma $L_C = 15$ entonces $N_C = L_C/N_F = 15/3 = 5$. Escribimos el criptograma en cinco columnas y luego leemos de arriba hacia abajo:*

| | | | | |
|---|---|---|---|---|
| M | A | P | D | D |
| I | T | O | O | E |
| R | U | R | N | X |

Obteniendo el siguiente mensaje $M = \text{MIRA TÚ POR DÓNDE}$.

Para hacer las cosas un poco más complicadas, otra forma de cifrar el mensaje, similar a la anterior, es mediante una figura de *zig-zag* de forma que la clave también se encuentra en el nivel de profundidad de dicha figura, como se indica en el próximo ejemplo.

Ejemplo 1.66: *Utilizando el cifrado por líneas con figura zig-zag con una profundidad igual a 3, cifre el mensaje $M = \text{EL ESPAÑOL COMO EL JUDÍO, DESPUÉS DE COMER SIENDE FRÍO}$.*

Solución: *El mensaje se escribe como se indica:*

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| E | P | L | O | U | D | U | E | E | E | F |
| L | S | A | O | C | M | E | J | D | O | E |
| E | Ñ | O | L | I | S | S | O | S | T | I |

Luego, leyendo en filas, se obtiene el criptograma:

$C = \text{E P L O U D U E E E F L S A O C M E J D O E P E D C M R I N E R O E Ñ O L I S S O S T I}$.

1.8.5. Criptoanálisis de los cifrados por transposición

La técnica de anagramación consistirá en la elección de un conjunto de elementos de una columna o fila, llamado *ventana*, y su posterior comparación con otras cadenas de caracteres en columnas o filas de igual longitud con el objeto de encontrar digramas comunes que han sido rotos por la transposición. La idea es que dicha ventana *recorre* todo el texto cifrado y en algún lugar coincidirán todos los digramas con los del mensaje original. Veamos el caso particular de un ataque a una cifra por columnas. Los pasos a seguir ante un cifrado que se sospeche sea de columnas, serán los siguientes:

- Calcular primero la distribución de frecuencia de los caracteres del criptograma. Si dicha distribución resulta similar a la característica del lenguaje, es muy posible que el criptograma en cuestión se corresponda con un cifrado por transposición.
- Se elige una cadena de al menos 7 caracteres del comienzo del criptograma y que se denominará *ventana*. Con esta ventana se recorrerá el resto del criptograma avanzando de incrementos de un carácter y en cada paso se compararán los digramas que aparecen, fruto de los caracteres de dicha ventana y del resto del texto cifrado. Aunque es recomendable la elección de una ventana grande para poder aplicar con lógica las estadísticas del lenguaje, el tamaño de dicha ventana deberá ser menor que el número de filas que resultase en una cifra por columnas o bien el número de columnas de una cifra por filas.
- Si la mayoría de los digramas presentan una alta frecuencia, esto indica que puede tratarse de dos columnas de la operación de cifrado. De esta forma, podemos reconstruir la matriz y, por tanto, describir el criptograma.

En resumen, la idea es que, dado que tras la operación de cifra se conservan todos los caracteres del texto en claro, eso sí permutados, al comparar un bloque que

será parte de una columna con otros bloques, es posible encontrar digramas de alta frecuencia y esto permitirá encontrar el período y, por tanto, romper el cifrado. Veamos cómo funciona este método a través de un ejemplo. Cifremos, por ejemplo, en cuatro columnas el siguiente mensaje $M = \text{ESTO NO HAY QUIEN LO ARREGLE}$.

| | | | |
|---|---|---|---|
| E | S | T | O |
| N | O | H | A |
| Y | Q | U | I |
| E | N | L | O |
| A | R | R | E |
| G | L | E | X |

Leyendo por columnas y agrupando en bloques de cinco caracteres obtenemos el criptograma $C = \text{ENYEA GSOQN RLTHU LREOA IOEX}$.

La separación de los digramas del texto en claro dentro del criptograma es, precisamente, el número de filas obtenidas al confeccionar la matriz. Luego, si elegimos una ventana, por ejemplo, igual a 4 caracteres ENYE , y la comparamos con los restantes bloques de 4 caracteres del criptograma, en algún momento se realizará la comparación de la ventana ENYE con los caracteres desplazados un período, es decir la cadena SOQN , obteniéndose en este momento los digramas ES , NO , YQ y EN . Puesto que de estos cuatro digramas ES , NO y EN son muy frecuentes en el lenguaje castellano, podríamos suponer que el período de la cifra viene dado por la distancia que hay desde el primer carácter de la ventana hasta el primer carácter de la cadena analizada, en este ejemplo los seis espacios que separan la E de la S en la palabra ESTO del mensaje. Luego, si escribimos el criptograma en seis filas, se llega a la matriz anterior que permite encontrar el mensaje original.

Podemos generalizar este método diciendo que se elige una ventana de un tamaño V caracteres, es decir:

$$\text{Ventana} = C_1 C_2 \dots C_V \quad 1.70$$

A continuación, se observan los digramas que se forman al recorrer con esta ventana el resto del texto, es decir, formamos los siguientes digramas:

$$\begin{aligned} &C_1 C_{V+1}, C_2 C_{V+2}, \dots, C_V C_{2V} \\ &C_1 C_{V+2}, C_2 C_{V+3}, \dots, C_V C_{2V+1} \\ &C_1 C_{V+3}, C_2 C_{V+4}, \dots, C_V C_{2V+2} \\ &\dots \end{aligned} \quad 1.71$$

En cada comparación de la ventana con un bloque, buscamos la frecuencia relativa de los digramas encontrados de acuerdo a la Tabla de Digramas del Anexo y se calculan la *media* de la muestra y la *desviación estándar*. Si la media es un valor alto y la desviación estándar es baja, quiere decir que todos los valores de $C_a C_b$ tienen alta probabilidad de ser parte de un texto en claro y que, además, la media alta no es debido solamente a algún digrama aislado de muy alta frecuencia. Si se dan estas condiciones entonces es probable que el período L del cifrado, las filas de la matriz en

el caso de una cifra por columnas, sea igual a la distancia en caracteres que separa a ambas cadenas p_{Cx} y p_{C1} , luego:

$$L = p_{Cx} - p_{C1} \quad 1.72$$

donde p_{Cx} es la posición relativa donde comienza la cadena que se está comparando y p_{C1} es la posición de inicio del criptograma y de la ventana. La media se calcula sumando las frecuencias relativas f_r de los digramas en el lenguaje:

$$\bar{X} = \frac{1}{V} \sum_{i=1}^V f_r \quad 1.73$$

siendo V el tamaño de la ventana en caracteres. La desviación estándar σ será:

$$\sigma = \sqrt{\frac{\sum_{i=1}^V (f_r - \bar{X})^2}{V}} \quad 1.74$$

Encontrado un período L , podemos intentar extender el tamaño de la ventana hasta dicho valor, con la idea de tener una cadena de caracteres igual a la de una columna en la operación de cifra, o por el contrario buscar L posiciones más adelante otra cadena para ver si también presenta digramas comunes con la que le precede.

Ejemplo 1.67: *Realice un ataque por anagramación sobre el criptograma que se indica eligiendo una ventana de tamaño $V = 4$*

$C = \text{TPNOT OAPO DRYAD OAURO SUNAS.}$

Solución: *Como $V = 4$ el bloque será TPNO que se comparará con TOAO, OAOP, AOPO, OPOD, etc. Usaremos la Tabla de Digramas del Anexo.*

| | | | | | | | | | |
|-------------|---|----|--------|---|----|-------|---|----|--------|
| T | t | Tt | 11 | o | To | 285 | a | Ta | 436 |
| P | o | Po | 225 | a | Pa | 181 | o | Po | 225 |
| N | a | Na | 332 | o | No | 222 | p | Np | 49 |
| O | o | Oo | 40 | p | Op | 131 | o | Oo | 40 |
| Media: | | | 152 | | | 205 | | | 188 |
| Desviación: | | | 136,77 | | | 56,44 | | | 161,32 |

Con la cadena OAOP de la segunda comparación, los digramas to, pa, no y op muestran una media alta asociada con una baja desviación estándar respecto a esa media. Asimismo, en la tercera comparación -cadena AOPO- aparece el digrama ta, de muy alta frecuencia, pero su efecto se enmascara con los digramas np y oo de baja frecuencia, dando una desviación alta. De lo anterior podríamos deducir que el período es igual a 5, el tamaño de la cadena más el número de comparaciones hechas antes de dar con el bloque de media alta y desviación baja.

Siguiendo con el método y suponiendo que el período es igual a 5, podríamos observar los digramas que aparecen desplazándonos en el criptograma 5 espacios; es decir, comparar por ejemplo la cadena encontrada OAOP con DRYA, luego la cadena DRYA con OAUR y finalmente la cadena OAUR con SUNA, es decir:

| | | |
|--------|--------|--------|
| ar 493 | ra 520 | au 72 |
| oy 35 | yu 8 | un 338 |
| pa 181 | ar 493 | ra 520 |

Puesto que se mantiene una media alta, se confirma que el período podría ser igual a 5. No obstante, en estas comparaciones no se cumple que la media alta vaya acompañada de una desviación baja; esto ocurre cuando el tamaño de la ventana es de sólo algunos caracteres como es en este caso y que hace *muy difícil y arriesgado* aplicar estadísticas alegremente. En este ejemplo como la clave $N_C = 5$ y el mensaje tenía muy pocos caracteres se obtienen pocas filas (de hecho $N_F = 5$) y por lo tanto nos ha forzado la elección de una ventana pequeña. Recuerde que éste es un método estadístico, por tanto no infalible, y que para tener un mínimo grado de confianza en los resultados será necesario contar con un criptograma de gran longitud y no pocas veces algo de intuición y suerte. Como ejercicio, descifre Ud. mismo este *noble mensaje*.

1.9. OTRAS TRANSFORMACIONES

Además de los métodos clásicos analizados en el capítulo, existen infinidad de algoritmos de cifrado más o menos ingeniosos. A continuación presentaremos algunas transformaciones aritméticas fáciles de implementar en un ordenador.

1.9.1. Transformación por adición

Puesto que la suma y la resta son operaciones que cuentan con inversa, se pueden utilizar como funciones de cifra. De hecho, este principio ya ha sido utilizado en los cifradores por sustitución. Si existe una correspondencia entre los caracteres a_i del alfabeto del mensaje con los dígitos c_j del alfabeto de cifrado con $0 \leq j \leq n-1$, serán posibles las dos operaciones de cifra que se indican:

$$C = E_K(M) = M + K \quad 1.75$$

$$C = E_K(M) = M - K \quad 1.76$$

En este tipo de cifrado, tomamos un bloque de n caracteres y la cifra será el valor numérico resultante de la adición o sustracción de dicho número con el que le corresponda a los n caracteres de la clave que coinciden con el bloque. Estos números luego pueden transmitirse o almacenarse en formato binario.

Por ejemplo, para cifrar el mensaje $M = LA MEJOR DEFENSA ES UN BUEN ATAQUE$ con la clave $K = NAPOLEÓN$ y una transformación por adición con bloques de tres caracteres, procedemos de la siguiente forma:

| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| M | = | L | A | M | E | J | O | R | D | E | F | E | N | S | A | E | S | U | N | B | U | E | N | A | T | A | Q | U | E |
| C | = | N | A | P | O | L | E | O | N | N | A | P | O | L | E | O | N | N | A | P | O | L | E | O | N | N | A | P | O |

La representación numérica será:

M = 11 00 12 04 09 15 18 03 04 05 04 13 19 00 04 19 21 13 01 21 04 13 00 20 00 17 21 04

K = 13 00 16 15 11 04 15 13 13 00 16 15 11 04 15 13 13 00 16 15 11 04 15 13 13 00 16 15
 C = 24 00 28 19 20 19 33 16 17 05 20 18 30 04 19 32 34 13 17 36 15 17 15 33 13 17 37 19

Si tomamos como bloque la cifra de cada tres caracteres, el criptograma final será los números:

C = 240.028 192.019 331.617 52.028 300.419 323.413 173.615 171.533 131.737 19.

Para descifrar basta con realizar la operación inversa. Esto es, el primero y el segundo bloques del criptograma C_1 y C_2 se descifrarían como sigue:

| | | | |
|---------|----------------|---------|----------------|
| C_1 | 240.028 | C_2 | 192.019 |
| $- K_1$ | <u>130.016</u> | $- K_2$ | <u>151.104</u> |
| M_1 | 110.012 = LAM | M_2 | 040.915 = EJO |

De la misma manera, podríamos cifrar aplicando una multiplicación del bloque del texto en claro por el bloque de la secuencia de una clave. No obstante, deberíamos representar los caracteres del alfabeto con dígitos $C_j > 0$ pues de lo contrario la operación inversa sería irrecuperable para texto en claro o clave igual a cero. Esto quiere decir que los caracteres deberán representarse desde 1 a 27 a diferencia de cómo veníamos haciéndolo $[0, 26]$ que era una imposición de la aritmética modular, que no es aquí el caso. Siguiendo entonces el mismo ejemplo anterior, y tomando ahora como bloques a digramas del mensaje, el primer criptograma será el producto del mensaje *LA* (12 01) cifrado con la secuencia de clave *NA* (14 01); luego:

$$C_1 = M_1 * K_1 = LA * NA = 1.201 * 1.401 = 1.682.601.$$

El problema que se manifiesta en este tipo de cifra (y en menor medida en la suma) es que el criptograma puede expandir la longitud del mensaje, requiriéndose más dígitos o bits para su representación. El caso contrario, una operación de cifra que reduzca la longitud del mensaje, sería la división, pero ahora el problema es que el resultado de tal operación no arroje un entero y, consecuentemente, nos sea imposible la recuperación de la información.

1.9.2. Transformación por conversión de base

Si en la representación del mensaje se produce un cambio de base, se obtiene un criptograma en el que la seguridad reside únicamente en el secreto de la base utilizada. Esta operación de cambio de base es válida para cifrar puesto que cuenta con inversa; es decir, si un mensaje se representa en base decimal y se cifra en bloques de un tamaño determinado convirtiendo cada bloque a base octal, por ejemplo, se podrá descifrar posteriormente dicho criptograma sencillamente convirtiendo de nuevo el número del sistema octal al decimal.

Al igual que en los métodos aritméticos anteriores, la longitud del mensaje podrá expandirse o reducirse, según sea el sistema de numeración que elijamos para cifrar en comparación con la base original. Por ejemplo, si el mensaje a cifrar es $M = SOS$, de acuerdo a los dígitos asignados en castellano a estos caracteres del 1 al 27, su representación decimal será: $SOS = 201620_{10}$. Si la operación de cifra consiste en

representar bloques de trigramas en el sistema octal, se tiene que $M = \text{SOS} = 201.620_{10}$ que se representa como $C = 611624_8$.

Ejemplo 1.68: *Comente las operaciones realizadas para obtener el criptograma anterior y demostrar que al convertir de base se recupera el mensaje.*

Solución: *Hacemos divisiones sucesivas del dividendo y cocientes por el divisor 8 hasta que el cociente sea menor que el divisor. Anímese y hágalo. Los restos de esta división, leídos en sentido inverso a partir del cociente distinto de la base entregan el resultado buscado. Resuélvalo Ud. mismo y compruebe que $C = 611624_8$.*

Para recuperar el mensaje, se convierte a formato decimal, es decir:

$$M = 4 \cdot 8^0 + 2 \cdot 8^1 + 6 \cdot 8^2 + 1 \cdot 8^3 + 1 \cdot 8^4 + 6 \cdot 8^5 = 201.620_{10}$$

1.9.3. Transformación por lógica de Boole

Utilizando el álgebra de Boole también podemos cifrar un mensaje. Una posibilidad sería usar las operaciones de negación y OR exclusivo que tienen inversos. Si el texto en claro se representa en binario conjuntamente con una clave también binaria de la misma longitud, esto dará lugar a las operaciones de cifrado y descifrado que se indican:

$$\text{Negación: } C = E(M) = \overline{M} \quad \text{y} \quad M = \overline{E(M)} \quad \boxed{1.77}$$

$$\text{XOR} \quad C = E_K(M) = M \oplus K \quad \text{y} \quad M = E_K(M) \oplus K \quad \boxed{1.78}$$

Ejemplo 1.69: *Se desea cifrar el mensaje $M = \text{SOL}$ representado en su equivalente ASCII binario: a) Cífrelo con el algoritmo de negación y b) Cífrelo con la función XOR y la clave $K = \text{SUN}$. En cada caso, represente el resultado en notación hexadecimal.*

Solución: *Las representaciones binarias del mensaje y la clave son:*

$\text{SOL} = 01010011 \ 01001111 \ 01001100$

$\text{SUN} = 01010011 \ 01010101 \ 01001110$

a) $C_1 = 10101100 \ 10110000 \ 10110011 = (\text{AC B0 B3})_{16}$

b) $C_2 = 00000000 \ 00011010 \ 00000010 = (\text{00 1A 02})_{16}$

En algunos casos como en el criptograma C_1 es posible representar la solución con caracteres ASCII imprimibles. Esto no será siempre posible como es lógico pues la operación puede dar lugar a octetos que no tengan representación impresa como sucede con todos los elementos del segundo criptograma. Ahora bien, esto no tiene mayor importancia pues la información se cifra, envía y descifra o en su caso se almacena en binario con lo que, evidentemente, se recupera el texto en claro usando la operación inversa; en el primer caso negando uno a uno los bits del criptograma y en el segundo aplicando nuevamente Or exclusivo al criptograma con la misma clave. Vea las tablas de caracteres en el Anexo y compruébelo.

1.9.4. Transformación matricial

Podemos generalizar los cifrados matriciales de Hill vistos en el apartado anterior, utilizando las operaciones producto y suma de matrices, sobre un mensaje M que se transforma mediante un código binario en una sucesión de unos y ceros que se disponen en una matriz de r filas por s columnas, al igual que la clave K . Luego, las operaciones de cifra serán:

$$(C) = (M) + (K) \quad 1.79$$

$$(C) = (M) \times (K) \quad 1.80$$

Para que exista inversa en la suma, será necesario que las matrices sean de la misma dimensión, en tanto que para el producto como ya hemos visto se debe cumplir que la matriz clave sea cuadrada, no singular y posea una inversa única.

Ejemplo 1.70: *Cifre el mensaje $M = \text{CAMISA}$, usando una transformación de suma de matrices como la indicada en la ecuación 1.79 con la clave $K = \text{ROSADA}$. La cifra es módulo 27 y se usarán matrices de 2 filas y 3 columnas.*

Solución: *Las matrices de texto y su correspondiente código mod 27 serán:*

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{pmatrix} = \begin{pmatrix} C & A & M \\ I & S & A \end{pmatrix} + \begin{pmatrix} R & O & S \\ A & D & A \end{pmatrix} = \begin{pmatrix} 2 & 0 & 12 \\ 8 & 19 & 0 \end{pmatrix} + \begin{pmatrix} 18 & 15 & 19 \\ 0 & 3 & 0 \end{pmatrix}$$

Procediendo a la suma módulo 27:

$$\begin{pmatrix} C_{11} & C_{12} & C_{13} \\ C_{21} & C_{22} & C_{23} \end{pmatrix} = \begin{pmatrix} 20 & 15 & 31 \\ 8 & 22 & 0 \end{pmatrix} \bmod 27$$

$C = 20 \ 15 \ 04 \ 08 \ 22 \ 00 = \text{TOD IVA}$

Observe la similitud con el cifrado de Vigenère.

SUMARIO DEL CAPÍTULO

1. La seguridad unida al secreto de un algoritmo de cifrado está relacionada con sistemas criptográficos clásicos. Los sistemas modernos basan su fortaleza en el secreto de la o las claves. En este caso se habla de algoritmos *simétricos* que usan clave secreta o algoritmos *asimétricos* que usan una única clave pública y otra privada.
2. Los cifradores clásicos se clasifican en Cifradores por *Sustitución* y Cifradores por *Transposición*. Entre los primeros están los monoalfabéticos y los polialfabéticos. En la misma categoría están a los cifradores por homófonos.
3. Los cifrados por *sustitución* aplican el principio de la *confusión* propuesta por Shannon, que consiste en sustituir caracteres del texto en claro por otros caracteres del mismo alfabeto o de otros alfabetos. Por su parte, los cifrados por *transposición* aplican el principio de la *dispersión* también propuesto por Shannon y que tiene como acción la permutación de los caracteres del texto en claro.
4. La *sustitución* permitirá aplicar distintos alfabetos y, por tanto, *enmascarar* las

distribuciones de frecuencias características de los caracteres del texto en claro. Su efecto es producir una relación texto en claro/criptograma lo más compleja posible. Por su parte, la *permutación* utiliza el mismo alfabeto del texto en claro o un equivalente de alfabeto mixto para el cifrado, provocando así una dispersión de los caracteres. En este caso, se conserva la distribución característica del lenguaje en el criptograma.

5. Los cifradores por sustitución simple monoalfabeto utilizan $E(m) = (a * m \pm b) \bmod n$ como transformación afín, siendo a la constante de decimación, m el valor numérico asociado al carácter del alfabeto, b el desplazamiento en dicho alfabeto y n el número de letras del alfabeto o cuerpo de trabajo. En general, cuando $a = 1$, diremos que se trata de un cifrador por *alfabetos desplazados puros* y si $b = 0$, de un cifrador por multiplicación o *decimación pura*.
6. El criptoanálisis de sistemas con cifrado por sustitución monoalfabética consiste en encontrar *correspondencias* entre caracteres en claro con los del criptograma, usando para ello las tablas de frecuencia de los monogramas. Se pretende formar trozos de palabras muy frecuentes aprovechando la *redundancia* del lenguaje.
7. Los cifradores por *homófonos* utilizan distintos valores para representar un mismo carácter, con el objeto de destruir la distribución de frecuencia de los caracteres del lenguaje. La fortaleza del cifrado está en la clave o el archivo que determina los homófonos. Para cifradores por homófonos de orden $n > 1$, se obtienen n criptogramas válidos en el espacio de mensajes M al descifrar o intentar describir el texto cifrado, dificultando de esta forma el ataque.
8. Los cifradores polialfabéticos utilizan más de un alfabeto para el cifrado. En este caso, existe una periodicidad d con la que se repite la misma transformación. En su ataque, para encontrar este período contamos con el *Método de Kasiski* y el Índice de Coincidencia.
9. El *Método de Kasiski* ayuda a determinar el período de un criptograma mediante la simple inspección de éste. Consiste en buscar *cadenas de caracteres repetidos* en el texto cifrado. El valor del máximo común divisor de las distancias entre tales cadenas será el período buscado.
10. El Índice de Coincidencia nos ayuda a determinar si los d textos cifrados independientes provenientes del criptograma se corresponden, cada uno, a sustituciones simples comparando su valor con el característico del lenguaje. Ambos métodos, Kasiski e Índice de Coincidencia, se complementan.
11. Los sistemas de cifra polialfabética con clave continua en la que dicha clave corresponde a un texto en el mismo lenguaje, suaviza la distribución de frecuencias. Si bien no existe período en este caso, puede criptoanalizarse mediante el *Método de Friedman* que consiste en buscar pares $M_i K_i$ de alta frecuencia que entreguen el carácter del criptograma analizado. Mediante combinaciones múltiples de texto en claro y claves posibles puede determinarse,

bajo ciertas condiciones, trozos del mensaje y de la clave en cuestión.

12. Los cifradores por sustitución *poligrámica* aplican las técnicas de sustitución sobre digramas, trigramas, etc., con el objeto de destruir la distribución de frecuencia característica de los monogramas. Son típicos el cifrador de Playfair y el de Hill.
13. El ataque al sistema de cifra poligrámico de *Playfair* se realiza mediante la técnica de *Análisis de Digramas* con el objeto de reconstruir la matriz de cifra.
14. El cifrado de *Hill* introduce el uso de las matemáticas de matrices. Aunque es un método de cifra muy interesante, no pudo competir con otras máquinas de cifrar de la época como el Enigma.
15. Aunque se obtienen con el sistema de Hill valores de distancia de unicidad muy altos y por lo tanto el ataque por fuerza bruta es inviable, presenta una gran debilidad ante ataques con texto en claro conocido.
16. Para atacar un cifrado de Hill conociendo el texto en claro, buscamos *vectores unitarios* en el texto en claro o en el criptograma, lo que nos dará de forma inmediata la matriz de la clave. Si no es posible encontrar dichos vectores unitarios, usamos la técnica de *Gauss-Jordan* que consiste en diagonalizar, por ejemplo, la matriz Texto Claro/Texto Cifrado, de lo cual puede deducirse la clave.
17. Los cifradores por transposición utilizan la *permutación de los mismos caracteres* del texto en claro realizando permutaciones de alfabetos, por grupos, columnas, filas y series, pudiendo utilizarse si se desea una clave. Este cifrado rompe las cadenas características del lenguaje.
18. El criptoanálisis de cifrados por transposición se realiza aplicando *Técnicas de Anagramación*, que consiste en la distribución del criptograma en bloques de digramas y la comprobación posterior, mediante una *ventana* de texto cifrado, de la aparición de digramas característicos del lenguaje con el uso de una tabla de digramas característicos. Si dentro de la ventana de comparación, la media de los valores de digramas que se forman es alta y la desviación baja, esto puede indicar la presencia de un período de cifra.
19. El cifrado de *Vernam* usa como clave una cadena de bits para realizar la *suma or exclusivo* con el mensaje. Aunque fue diseñado para cifrar mensajes representados por el código Baudot, el cifrador de Vernam puede usarse para cifrar de forma binaria cualquier archivo, con o sin formato, mediante la representación en código ASCII o ANSI de todos los caracteres.
20. Si la secuencia de clave de un cifrador de Vernam binario es de tipo aleatorio y además de uso único, el sistema de cifra conocido como *one-time pad* es perfecto en tanto resulta imposible deducir la clave. Esta idea dará luego paso a los denominados cifradores de flujo.
21. Otros tipos de cifrado son las transformaciones por adición y sustracción de texto

en claro con claves, puesto que esta operación tiene inversa. También es posible la transformación por conversión de la base del sistema de numeración, transformaciones mediante operaciones lógicas *booleanas* y transformaciones matriciales en general.

22. Las máquinas de cifrar corresponden a sistemas implantados en equipos electromecánicos entre los que cabe destacar las máquinas a rotor como el *Enigma* y *Hagelin*, utilizadas para cifrar mensajes durante la 2ª Guerra Mundial.
23. Otras máquinas de cifra anteriores al siglo XX que conforman los pilares de la criptografía clásica, son el cifrador polialfabético de *Alberti* y los de *Wheatstone* y *Bazeries*, estos últimos basados en discos.