Actividad 4: Herramientas para criptografía y criptoanálisis

Seguridad Informática

Orquidea Seijas
ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA Grado en Ingeniería Informática

TABLA DE CONTENIDO

1. ATAQUE AL CIFRADO DE SUSTITUCIÓN MONOALFABÉTICO

a. INTRODUCCIÓN SOBRE EL CIFRADO DE SUSTITUCIÓN MONOALFABÉTICO.

La sustitución monoalfabética consiste en construir un alfabeto cifrado colocando al azar las letras del alfabeto plano. Así, es posible mejorar el cifrado del césar, que es posible descifrar mirando 27, ya que con este tipo de cifrado se pueden realizar *27!* alfabetos cifrados. El cifrado de sustitución monoalfabético constituye la familia de métodos criptográficos más simple de criptoanalizar, puesto que las propiedades estadísticas del texto plano se conservan en el criptograma.

b. JUSTIFICACIÓN TEÓRICA DE LA BASE DEL MÉTODO DE CRIPTOANÁLISIS UTILIZADO.

Para obtener el alfabeto plano, es posible aplicar ingeniería inversa sobre el alfabeto criptoanalizado. Por ejemplo, en inglés la letra más frecuente es la *e*, con un 13% de frecuencia relativa y la siguiente más frecuente, con bastante diferencia entre una y otra, es la *a*, con un 8%. Sabiendo esto, se podría desencriptar un texto empezando a buscar la letra con más frecuencia relativa y sustituyéndola por una e.

Para realizar este análisis, es necesario tener en cuenta la estadística de frecuencia del lenguaje en el que suponemos que el texto está escrito (inglés, castellano, etc.) y utilizar algún tipo de herramienta de análisis estadístico para estudiar el texto encriptado.

c. ETAPAS SEGUIDAS EN EL ANÁLISIS

En primer lugar, es necesario buscar similitudes entre las frecuencias de letras. La letra encriptada que más se repita probablemente será la más repetida en el idioma del texto sin encriptar. Para esta práctica, se intercambió un texto encriptado con un compañero de clase en inglés. Por esta razón, para este caso concreto, la letra más frecuente es la *e* y, en el texto encriptado, la *q*, con un 14%, tal y como se puede ver en la *Ilustración 1*.

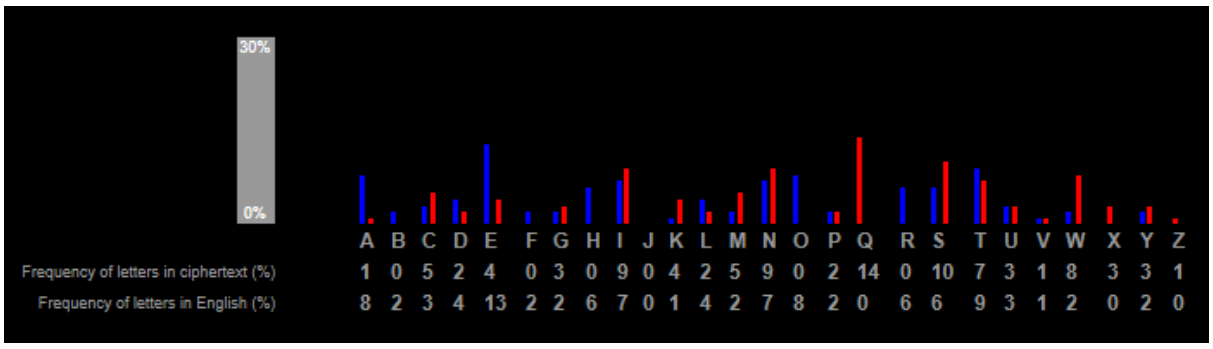


Ilustración 1 Frecuencia de letras en el texto cifrado y en inglés

A continuación, se decidió observar las palabras más cortas que había en el texto y que contenían la *e*. En inglés, una palabra comúnmente utilizada es el artículo *the*. Así pues, se observó la frecuencia de las letras cifradas que se presuponía que debían corresponder con la *t* y la *h*. Ambas estaban lo suficientemente cerca como para ser consideradas como posibles, por lo que se sustituyeron con el fin de comprobar si tenía sentido su posición en el texto. Puesto que solo había tres letras desencriptadas, no era posible comprobar por el momento si el encriptado se estaba realizando de manera apropiada.

Finalmente, se decidió seguir comparando letras en búsqueda de frecuencias similares con el fin de encontrar más letras para resolver el texto cifrado. También se decidió utilizar conocimientos sobre el idioma para completar palabras una vez faltara una o dos letras y fuera evidente la palabra a descifrar. Fue posible obtener todas las frecuencias de las letras que aparecían en el texto. Sin embargo, las letras *j* y *k* no aparecían en el texto y, por lo tanto, no se encontró su clave cifrada.

d. CASO PRÁCTICO

i. TEXTO CIFRADO A ATACAR

El texto encriptado era el siguiente:

CTNSKQM NG SKQ OQTQGWSI NG SKQIQ TQV DCUDQSI WI SKQ KPDQ IXQQU WTEMQZQTS GNM MQZNSQ ENZZPTWECSWNTI. WT UCLI NG LNMQ, SVN XQMINTI CWZWTD SVN QFEKCTDQ WTGNMZCSWNT TQQUQU SN VCWS UNRQTI NG UCLI SN DQS SKQ MQIXNTIQ NG SKQ ENPTSQMXCMS. SKWI WI QIXQEWCYYL MQYQACTS WT SKQ ENTSQFS NG IEWQTSWGWE UWCYNDPQI WTANYAWTD MQIQCMEKQI YNECSQU WT IWDTWGWECTSYL UWISCTS XCMSI NG SKQ VNMYU. GNM SKWI MQCINT, WS WI XNIIWOYQ SN ICL SKCS IEWQTEQ MQHPWMQI PIWTD SKQIQ SNNYI WT NMUQM SN UQAQYNX CS SKQ XCEQ WS WI EPMMQTSYL IPXXNIQU SN UN.

ii. TEXTO PLANO RECUPERADO

El texto recuperado, tras revisar la puntuación y las mayúsculas y minúsculas, era el siguiente:

Another of the benefits of these new gadgets is the huge speed increment for remote communications. in days of yore, two persons aiming two exchange information needed to wait dozens of days to get the response of the counterpart.Tthis is especially relevant in the context of scientific dialogues involving researches located in significantly distant parts of the world. for this reason, it is possible to say that science requires using these tools in order to develop at the pace it is currently supposed to do.

iii. CLAVE.

La clave obtenida es la siguiente:

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

2. ATAQUE AL CIFRADO DE SUSTITUCIÓN POLIALFABÉTICO (VIGENÈRE)

a. INTRODUCCIÓN SOBRE EL CIFRADO DE SUSTITUCIÓN MONOALFABÉTICO.

La sustitución polialfabética consiste en utilizar, como mucho, tantos alfabetos como letras tenga un idioma. Por lo tanto, en castellano se hablaría de la posibilidad de utilizar hasta 27 alfabetos mientras que en inglés sería posible utilizar hasta 26 alfabetos. De estos *x* alfabetos diferentes, se puede escoger cualquiera para cifrar cada letra del mensaje a encriptar. Este tipo de cifrado se consideraba *imposible de romper*, por la revista *Scientific America*, a principios del siglo XX, a pesar de que el método *Kasiki* resolvió el cifrado en el siglo XIX.

Para que este método de encriptación funcione, es necesario que haya una clave previamente compartida entre las dos partes involucradas. Esto es debido a que, en función de términos matemáticos, se puede expresar la función de cifrado como:

$$E(X_i) = (X_i + K_i)\ mod\ L$$

Donde X_i es la letra en la posición *i* del texto a cifrar, K_i es el carácter de la clave correspondiente a X_i, pues se encuentran en la misma posición y *L* es el tamaño del alfabeto. Para desencriptar, simplemente es necesario realizar la operación inversa (siempre teniendo en cuenta que si X_i -K_i es menor que cero, hay que sumar *L*).

Tal y como se puede ver, debido a la expresión utilizada, es posible que. a una misma letra en el texto plano, le puedan corresponder diferentes letras en el texto cifrado. Mientras más larga sea la palabra por utilizar, más fuerte será, por lo tanto, el cifrado polialfabético de descifrar. Así pues, ya no se puede aplicar directamente el método de la frecuencia estadística para descifrar el texto encriptado: la distribución de las letras es más uniforme y, por lo tanto, no es tan fácil observar qué letra cifrada corresponde a qué letra plana.

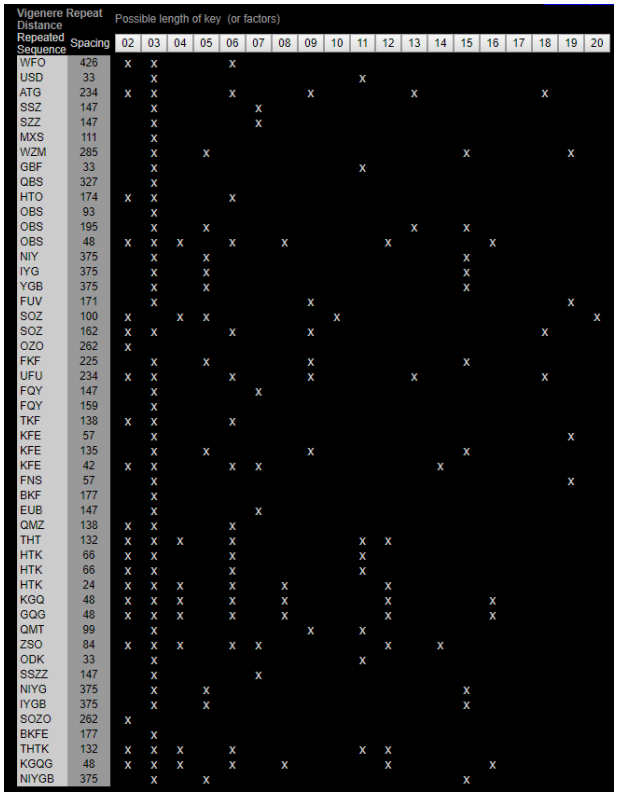
b. JUSTIFICACIÓN TEÓRICA DEL MÉTODO DE ATAQUE AL CIFRADO DE VIGENÈRE.

En este caso, el ataque al cifrado de Vigenère se basará, inicialmente, en encontrar la longitud de la clave utilizada. Para esto, será necesario analizar el texto en busca de repeticiones, esto es porque el cifrado polialfabético es un cifrado monoalfabético para cada una de las letras que funcionan como clave del mismo.

Una vez se sepa la longitud de la clave, se podrá dividir el texto en *n* grupos cifrados de forma monoalfabética. Estos grupos serán descifrables empleando el ataque utilizado en el apartado 1.b. que se basaba en conocimiento tanto del idioma del texto a atacar como de la frecuencia estadística de las letras del mismo.

c. ETAPAS SEGUIDAS EN EL ANÁLISIS.

En este caso, el texto a descifrar también se obtuvo de un compañero de clase. En primer lugar, se introdujo el texto plano, sin signos de puntuación o espacios, en la herramienta para atacar el cifrado de Vigenère de la web *The Black Chamber*. Se han buscado las secuencias repetidas con el fin de conocer la longitud de la clave para empezar a analizar el texto. Tal y como se puede ver en la *Ilustración 3*, la longitud más probable era de 3 caracteres.



Vigenere Repeat Distance Repeated Sequence	Spacing	Possible length of key (or factors)																	
		02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20
WFO	426	x	x			x														
USD	33		x								x									
ATG	234	x	x			x			x				x					x		
SSZ	147		x				x													
SZZ	147		x				x													
MXS	111		x																	
WZM	285		x		x										x				x	
GBF	33		x								x									
QBS	327		x																	
HTO	174	x	x			x														
OBS	93		x																	
OBS	195		x		x								x		x					
OBS	48	x	x	x		x		x				x				x				
NIY	375		x		x										x					
IYG	375		x		x										x					
YGB	375		x		x										x					
FUV	171		x						x										x	
SOZ	100	x		x	x					x										x
SOZ	162	x	x			x			x									x		
OZO	262	x																		
FKF	225		x		x				x						x					
UFU	234	x	x			x			x				x					x		
FQY	147		x				x													
FQY	159		x																	
TKF	138	x	x			x														
KFE	57		x																x	
KFE	135		x		x				x						x					
KFE	42	x	x			x	x							x						
FNS	57		x																x	
BKF	177		x																	
EUB	147		x				x													
QMZ	138	x	x			x														
THT	132	x	x	x		x					x	x								
HTK	66	x	x			x					x									
HTK	66	x	x			x					x									
HTK	24	x	x	x		x		x				x								
KGQ	48	x	x	x		x		x				x				x				
GQG	48	x	x	x		x		x				x				x				
QMT	99		x						x		x									
ZSO	84	x	x	x		x	x					x		x						
ODK	33		x								x									
SSZZ	147		x				x													
NIYG	375		x		x										x					
IYGB	375		x		x										x					
SOZO	262	x																		
BKFE	177		x																	
THTK	132	x	x	x		x					x	x								
KGQG	48	x	x	x		x		x				x				x				
NIYGB	375		x		x										x					

Ilustración 2 Uso de la herramienta The Black Chamber para observar la probabilidad de la longitud de la clave.

Puesto que se trata de una clave relativamente pequeña, de tres caracteres, las frecuencias no se ven tan alteradas como podrían verse y, por lo tanto, es posible establecer una correlación, al menos eso parece, entre la letra con más frecuencia del inglés, la e, y la letra con más frecuencia del primer alfabeto. Así pues, se puede inferir que la primera letra del primer alfabeto es la o, tal y como se pude ver en la *Ilustración 3.*

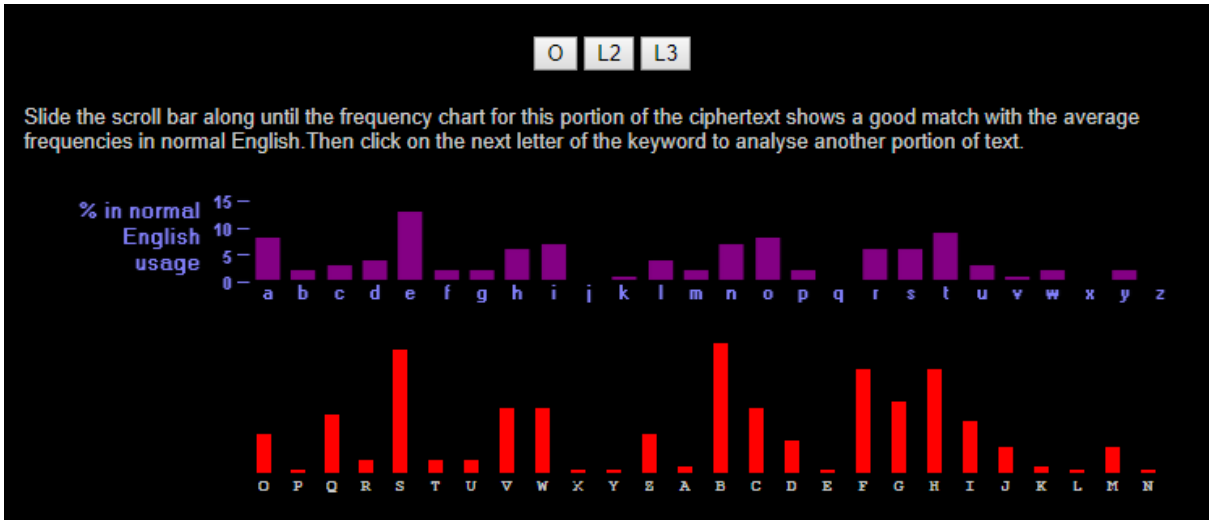


Ilustración 3: Frecuencia del primer alfabeto

A continuación, se analizaron los siguientes alfabetos, utilizando la misma suposición: las letras más frecuentes deben coincidir para obtener el alfabeto. Así, con la última letra, fue posible obtener un texto en inglés, correcto una vez se pusieron los signos de puntuación y espacios adecuados.

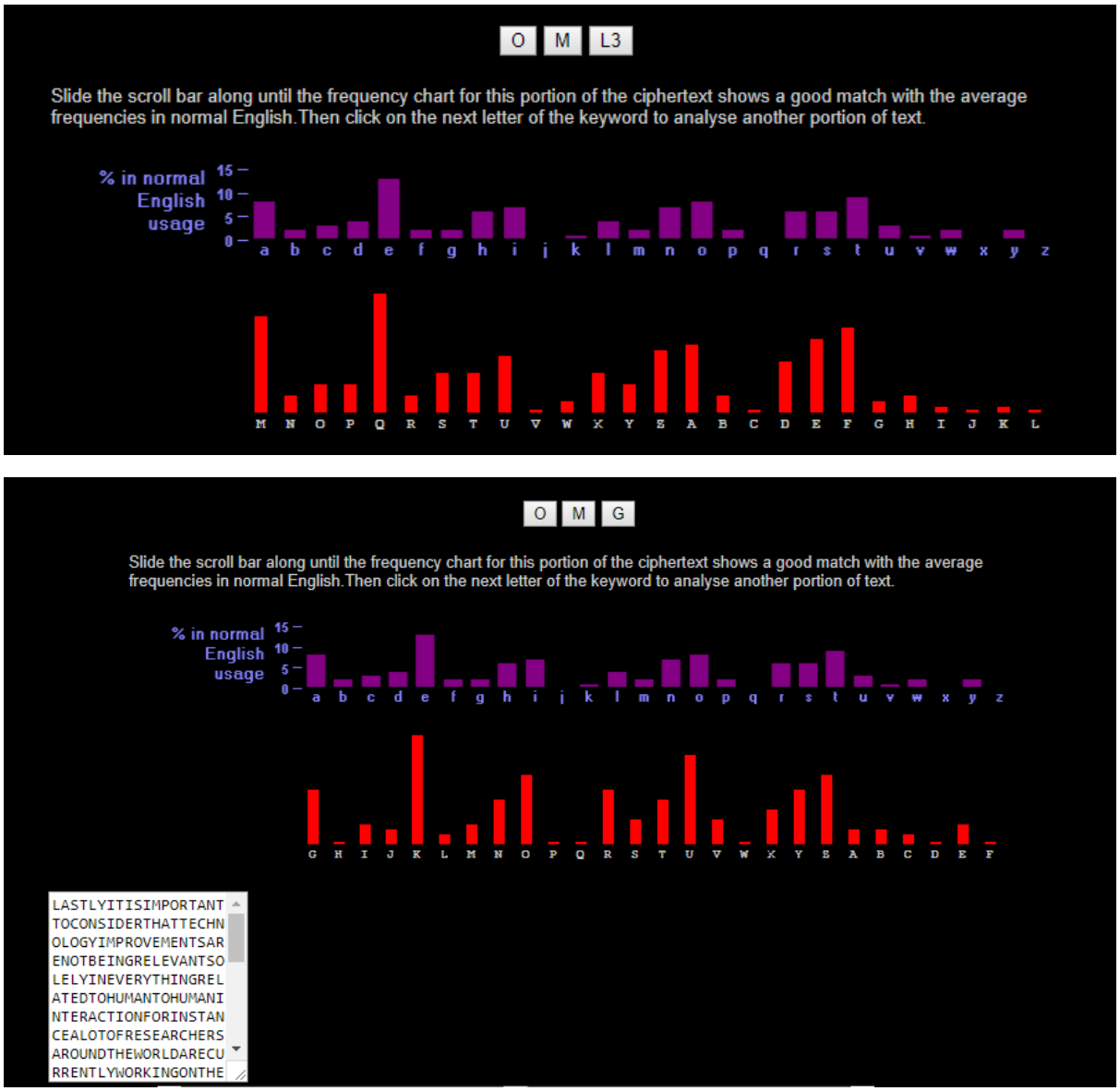


Ilustración 4 Tercer alfabeto y resolución

d. CASO PRÁCTICO

i. TEXTO CIFRADO A ATACAR.

Este fue el texto cifrado a utilizar:

ZMYHXEWFOGUSDAXHMTHFUQATGUJSDZVMZHQIVZUZAMMUSDDUJQSSZZGMXSZUHNKW
ZMFQRSHGBFYCXKZKOBQBSDEHTOBSXSXGHQJHANIYGBFUVGSOZOBFKFMIHUUBRUFUTGFGBOKOX
UHALFQYSMXQTKFEGFAABPZVQCCDRRMXSOAFDKBFRMIUFWOBSUBFNSPKJQRCBSSZZCRNMBKFEVS

OZFMRWYGUQGBMRMEOGMRUAXWFNAEUBQUTFNSMVDXOQMZWATGALHTOGEZIPETUKZPOGFU
VQRDATHTKFQYQGKCRYVUVKDKQWYIDBWHUFEZVUYWEVCEYWNRSNKQMAGQICYVIFKFEXIZTWZM
HTKGQGZSUFUZVYYQMTRQZSOZOEOBSRSBKFEUBXUQMZSPOBMNISKGQGODKOQBSZCVQTHTKZUM
VFOBSICZJWFOCZYODKHTKDAUFQYHFNIEZSONBARCSEQMTSHKBEGJQNIYGBXOJQY

ii. TEXTO PLANO RECUPERADO.

Este fue el texto recuperado:

LASTLY, IT IS IMPORTANT TO CONSIDER THAT TECHNOLOGY IMPROVEMENTS ARE NOT BEING RELEVANT SOLELY IN EVERYTHING RELATED TO HUMAN TO HUMAN INTERACTION. FOR INSTANCE, A LOT OF RESEARCHERS AROUND THE WORLD ARE CURRENTLY WORKING ON THE DEVELOPMENT OF HYPERSPECTRAL IMAGE ANALYSIS ALGORITHMS. ONE OF THE APPLICATIONS OF THIS STUDYFIELD, IS TO HELP ON THE RESCUE OF SHIPWRECK SURVIVORS. THIS IS POSSIBLE BECAUSE COMPUTERS RUNNING THESE ALGORITHMS CAN DETECT A SINGLE PERSON LOCATED IN A HUGE SEA AREA EVEN WHEN THE LIGHTING CONDITIONS ARE THE POOREST. THUS, TECHNOLOGY CAN EVEN SAVE HUMAN LIVES.

iii. CLAVE UTILIZADA.

La clave obtenida fue esta: OMG.

3. MÁQUINA DE CIFRADO ENIGMA

Enigma I es una máquina de cifrado electromecánica, desarrollada en 1927 para el ejército alemán. Esta máquina fue utilizada en la Segunda Guerra Mundial por el ejército de tierra y mar alemán.

a. JUSTIFICACIÓN SOBRE LA ROBUSTEZ DEL CIFRADO ENIGMA.

Enigma I tiene tres rotores de cifrado, seleccionados de un conjunto de cinco ruedas añadidas en 1938, cada una con 26 contactos a cada lado. Cada día, el operario de la máquina seleccionaba las tres ruedas indicadas en un orden concreto y guardaba el resto.

El uso de varios rotores permitió un modo simple de determinar qué alfabeto de sustitución usar para un mensaje en particular (en el proceso de cifrado) y para un texto cifrado (en el de descifrado). Este proceso es similar al proceso del cifrado polialfabético. Sin embargo, a diferencia de la mayoría de las variantes del sistema polialfabético, Enigma no tenía una longitud de clave fácilmente deducible. Los rotores generaban una nueva sustitución alfabética en cada pulsación: toda la secuencia de alfabetos de sustitución podía cambiarse haciendo girar uno o más rotores, cambiando el orden de los rotores, etc., antes de comenzar una nueva codificación. En el sentido más simple, Enigma tenía 26 x 26 x 26 = 17.576 alfabetos de sustitución para cualquier combinación y orden de rotores dada. Siempre y cuando el mensaje original no fuera de más de 17.576 pulsaciones no habría un uso repetido de un alfabeto de sustitución.

Enigma permitía una encriptación aún más robusta que la que se ha explicado en el párrafo anterior: la secuencia de los alfabetos utilizados era diferente si los rotores eran colocados en la posición ABC o en la posición ACB. Había un anillo que rotaba en cada rotor que se podía fijar en una posición diferente, y la posición inicial de cada rotor era también variable.

Al principio de cada mes, se les daba a los operadores de la Enigma un libro que contenía las configuraciones iniciales para la máquina. Como los rotores podían permutarse en la máquina, con tres rotores en tres posiciones se obtienen otras 3 x 2 x 1 = 6 combinaciones, para dar un total de 105.456 (17.576x6) posibles alfabetos.

La versión militar Enigma I contaba con una tabla de conexiones, *Steckerbrett*, en la que se podían intercambiar cualquier par de letras. Así pues, sabiendo que había 26 letras, había, por lo tanto, 26 entradas en la tabla de conexiones con un máximo de 13 conexiones entre las letras. Cualquier número de cables entre 0 y 13 era posible. El número total de combinaciones para n cables es el siguiente:

$$n = \frac{26!}{n! * (26 - 2n)! * 2^n}$$

Normalmente, el ejército alemán utilizaba diez cables para intercambiar 20 letras, por lo que el número de posibles combinaciones era de: 150.738.274.900.000.

b. Funcionamiento a ambos lados del canal de comunicación la clave de cifrado para el caso de la versión de Enigma I.

El operador de Enigma seleccionaba configuraciones para los rotores, definiendo las posiciones o "giros" de los rotores. Un operador en particular podría seleccionar ABC, y así se convertían en la configuración del mensaje para esa sesión de cifrado. Entonces se tecleaba la configuración del mensaje en la máquina que aún estaba con la configuración inicial. Los alemanes, creyendo que le otorgaban más seguridad al proceso, lo tecleaban dos veces. Sin embargo, tal y como se pudo observar con el cifrado de Vigenère, las repeticiones dan lugar a la posibilidad de notar un patrón y esto resultó clave para descifrar a Enigma.

Una vez tecleada la configuración y obtenido el cifrado asociado, el operador giraba los rotores a la configuración del mensaje, ABC, y procedía a escribir el resto el mensaje. En el extremo receptor, el funcionamiento se invertía. El operador ponía la máquina en la configuración inicial e introducía las primeras seis letras del mensaje. Al hacer esto él vería ABCABC en la máquina. Entonces el operador giraba los rotores a ABC e introducía el resto del mensaje cifrado, descifrándolo.

Este sistema era excelente porque el criptoanálisis se tendía a basar en análisis de frecuencias. Aunque se enviaran muchos mensajes en cualquier día con seis letras a partir de la configuración inicial, se asumía que esas letras eran al azar.

REFERENCIAS

- *Letter frecuency.* 2017. Wikipedia. https://en.wikipedia.org/wiki/Letter_frequency [Última visita: 15/10/17]
- *DMA.* 2017. http://www.dma.fi.upm.es/recursos/aplicaciones/matematica_discreta/web/aritmetica_modular/cesar.html [Última visita: 15/10/17]
- *Criptografía simétrica (I).* P. Cariñena. 2017.
- *Cifrado de Vigenère.* 2017. Wikipedia. https://es.wikipedia.org/wiki/Cifrado_de_Vigenère [Última visita: 15/10/17]
- *Why is Vigenère so strong?* 2017. The Black Chamber. http://www.simonsingh.net/The_Black_Chamber/vigenere_strength.html [Última visita: 15/10/17]
- *Enigma I.* Crypto Museum. http://www.cryptomuseum.com/crypto/enigma/i/index.htm [Última visita: 15/10/17]
- *Enigma (máquina).* 2017. Wikipedia. https://es.wikipedia.org/wiki/Enigma_(máquina) [Última visita: 15/10/17]