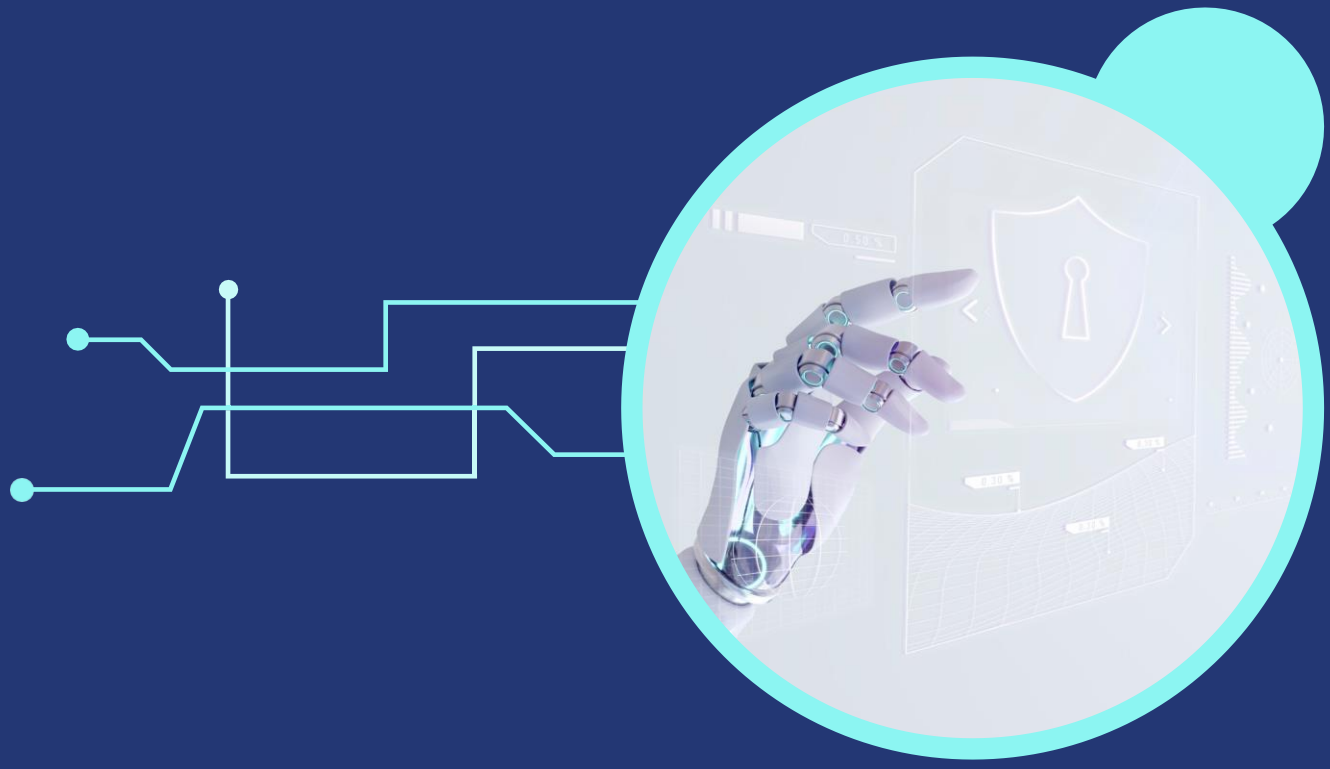
MÓDULO 1. CRIPTOGRAFÍA (UNIDAD 1)

PREMÁSTER EN

ANÁLISIS DE MALWARE, REVERSING Y BUG BOUNTY



Campus Internacional CIBERSEGURIDAD

ENIIT INNOVA IT BUSINESS SCHOOL

UCAM Universidad Católica de Murcia

INDICE DE CONTENIDOS

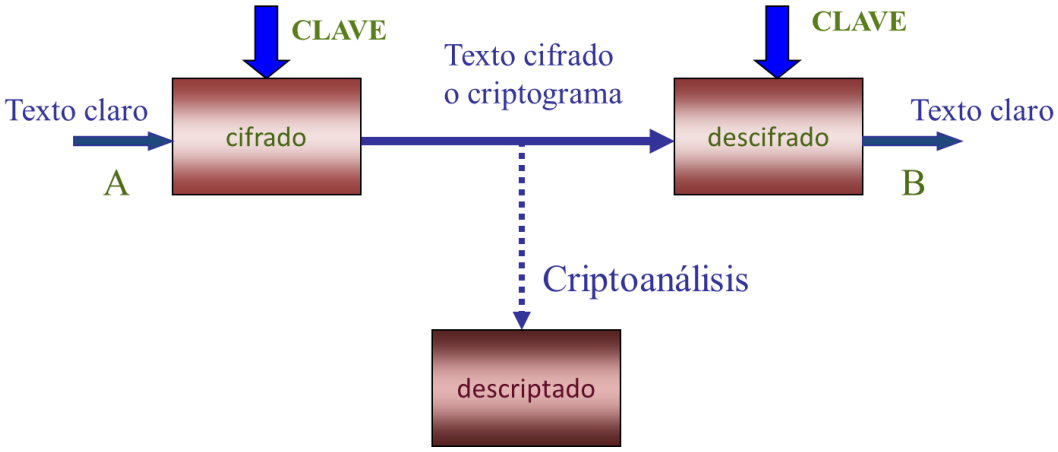
1. Introducción y algoritmos clásicos

El término *criptología* proviene del griego *krypto*: 'oculto' y *logos*: 'palabra'. Hace referencia a la ciencia de ocultar mensajes secretos y descubrirlos.

La criptología engloba a la criptografía y al criptoanálisis. La **criptografía** diseña algoritmos de cifrado para proteger la información. Como ciencia opuesta a la criptografía, el **criptoanálisis** se ocupa de romper dichos métodos de cifrado para recuperar la información original.

1.1. Conceptos iniciales

El procedimiento criptográfico funciona como puede verse en la figura.



En esta figura el emisor *A* quiere enviar un mensaje secreto al receptor B por un canal convencional, público y que podría ser inseguro. Para ello *A* cifra el mensaje utilizando la clave de cifrado, dando lugar al texto cifrado o criptograma, que envía por el canal de comunicación para que lo reciba *B*. Por su parte, *B* recibe el texto cifrado y haciendo uso de la clave puede descifrar el mensaje y obtener de nuevo el texto claro original del mensaje que envió *A*.

Si entre medias un atacante interceptara el mensaje cifrado y empleara técnicas de criptoanálisis, podría llegar a romper el cifrado y obtener el mensaje descriptado.

La criptografía está relacionada con la seguridad de la información, la cual abarca varios aspectos:

- Confidencialidad: consiste en garantizar que la información es accesible sólo para aquellos autorizados a tener acceso.
- Autenticidad del emisor y receptor: se trata de verificar que efectivamente se trata de las personas o entes que envían el mensaje o lo reciben.
- Integridad del criptograma, es decir, que éste no sufra ningún tipo de alteración o modificación
- No Repudio de origen: el emisor no puede negar que envió un mensaje porque el destinatario tiene pruebas del envío. En este caso la prueba la crea el propio emisor y la recibe el destinatario.
- No Repudio de destino: el receptor no puede negar que recibió el mensaje porque el emisor tiene pruebas de la recepción. Este servicio proporciona al emisor la prueba de que el destinatario legítimo de un envío realmente lo recibió, evitando que el receptor lo niegue posteriormente. En este caso la prueba irrefutable la crea el receptor y la recibe el emisor.

Una de las finalidades de la criptografía es mantener la confidencialidad del mensaje, esto es, garantizar que la información es accesible únicamente para aquellos autorizados a tener acceso. Otra es garantizar la integridad del criptograma, es decir, que éste no sufra ningún tipo de alteración o modificación. Así como garantizar la autenticidad del emisor y receptor, que se refiere a verificar que efectivamente se trata de las personas o entes que envían el mensaje o lo reciben.

1.2. Clasificación de los sistemas criptográficos

En función de las claves que se usen para cifrar y descifrar los mensajes, los sistemas criptográficos se pueden clasificar de la siguiente manera:

- Sistemas de **clave secreta**. En estos sistemas se utiliza la misma clave tanto para cifrar como para descifrar el mensaje. Por esta razón también se conocen como sistemas **simétricos**. La clave debe ser secreta y conocida tanto por el emisor como por el receptor, y únicamente por ellos. Este es un problema que no resulta tan sencillo de resolver. Para que estos

sistemas funcionen correctamente se asume que previamente ambos interlocutores se han puesto de acuerdo en la clave a utilizar o que han hecho uso de alguna metodología de distribución de claves que ha hecho llegar la clave de forma segura a ambos interlocutores.

- Sistemas de **clave pública**. En este caso se utiliza una clave diferente para cifrar y descifrar, por ello también se denominan sistemas **asimétricos**. Habitualmente ambos extremos tienen un par de claves, una pública y otra privada. En estos sistemas no es necesario un procedimiento de compartición de clave.

A su vez, dentro de los sistemas de clave secreta se puede distinguir entre diferentes tipos de sistemas:

- Cifrado en **flujo**. En este caso la transformación de cifrado se realiza sobre cada carácter del mensaje original. En los sistemas clásicos generalmente esta transformación se aplicaba sobre los caracteres del alfabeto en el que estaba escrito el mensaje, sin embargo, hoy en día se suele utilizar un alfabeto binario y las transformaciones se aplican sobre cada bit.
- Cifrado en **bloque**. A diferencia del cifrado en flujo, en el cifrado en bloque la transformación de cifrado se aplica sobre un grupo de caracteres del mensaje original.

1.3. Principios básicos

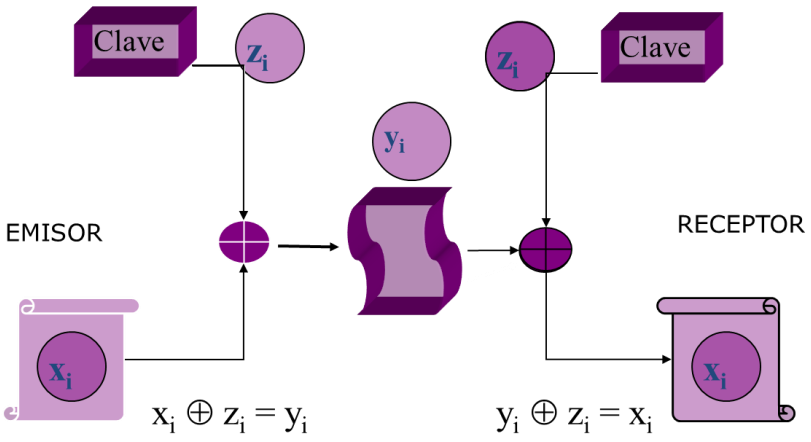
En la criptografía se han empleado habitualmente una serie de operaciones para realizar transformaciones en el texto y de esta forma cifrarlo para que fuera ininteligible para cualquier persona a la que no fuera dirigido el mensaje secreto.

Dos de estos principios básicos son la sustitución y la transposición.

1.3.1. Sustitución

La **sustitución** consiste en establecer una correspondencia entre el alfabeto en el que está escrito el mensaje y otro alfabeto, que no tiene necesariamente que ser el mismo que el alfabeto original. De esta manera, se van sustituyendo los caracteres del mensaje original por el correspondiente asignado en el alfabeto de salida. Para descifrar el mensaje el receptor debe conocer cuál es la correspondencia y hacer la sustitución inversa para recuperar el texto original.

El esquema de cifrado se puede ver en esta figura:



El emisor envía un mensaje x_i. Para obtener el criptograma y_i que enviará al receptor, aplica una operación XOR al mensaje y la clave z_i. Cuando el receptor recibe el criptograma y_i, realiza una operación XOR del criptograma recibido y_i junto con la clave (la misma que utilizó el emisor para cifrar), obteniendo el mensaje en claro x_i.

Un ejemplo de este tipo de cifrado es el famoso cifrado César, que veremos más adelante.

Una de las debilidades de la sustitución es que la frecuencia de aparición de cada letra en el mensaje original se refleja en el criptograma. Esto lo hace débil ante un ataque de análisis de frecuencias. Este tipo de ataque se basa en la frecuencia con la que aparecen los caracteres en los diferentes idiomas. De esta manera, conociendo las letras de mayor frecuencia del alfabeto utilizado se puede deducir la clave.

1.3.2.Transposición

La **transposición** consiste en barajar los símbolos del mensaje original y colocarlos en distinto orden con el objeto de hacer el mensaje ininteligible. En este caso los símbolos del criptograma son los mismos que los del mensaje original pero descolocados. Es necesario que el receptor conozca la transposición para recolocar los símbolos en el orden original y reconstruir el mensaje.

Un ejemplo de transposición es el siguiente:

Grupos de 4 letras

Transposición: 1234-> 4321

Mensaje	SISTEMAS	CLASICOS
Criptograma	TSISSAME	SALCSOCI

La clave secreta es el número de letras en cada grupo y el orden en el que se colocan.

La escítala lacedemonia, que veremos en el siguiente apartado, es un sistema de cifrado por transposición.

1.4. Algoritmos clásicos

Cabe mencionar que la Criptografía es tan antigua como la propia escritura. Resulta curioso que tan pronto como el ser humano aprendió a escribir surgió la necesidad de proteger determinada información y ser capaces de transmitir mensajes de forma secreta.

La historia de la criptografía es amplia y tiene muchos sistemas criptográficos y criptoanálisis a estudiar. En este apartado se hace un breve repaso a algunos de los sistemas más destacados. El repaso realiza un viaje desde la aparición de la criptografía hasta la criptografía científica, lo cual permite entender cómo han evolucionado los criptosistemas y se han ido haciendo más complejos con el paso del tiempo.

1.4.1. Inicios

Como precedentes a la criptografía se pueden mencionar la escritura jeroglífica, en particular en jeroglíficos no estándares tallados en monumentos del Antiguo Egipto. En un principio éstos no estaban previstos para ocultar la información sino para despertar el interés, conseguir misterio, realzar la figura del homenajeado, exaltación del escriba o incluso diversión para el espectador.

En relación a los jeroglíficos es de gran importancia la **piedra roseta**. Ésta data del año 196 a.C. y se puede ver en la siguiente figura.

Esta piedra contiene el mismo texto escrito con tres escrituras diferentes:

- Jeroglíficos egipcios.
- Escritura demótica (una variedad de la lengua egipcia).
- Griego antiguo.

Es precisamente esto lo que le da su importancia histórica, ya que ayudó a entender el significado de los jeroglíficos egipcios.

La **escítala lacedemonea** se considera como uno de los primeros sistemas criptográficos conocidos. Su nombre proviene del pueblo griego de Esparta, también conocido como Lacedemonia, que creó este sistema para proteger sus secretos. Este sistema data del S. V a.C. Consiste en un bastón en él se enrolla una cinta y en la cual se escribe longitudinalmente el mensaje secreto que se quiere transmitir, en los trozos visibles de la cinta. En la figura se puede ver un ejemplo de escítala.

Para transmitir el mensaje, la cinta se desenrolla y se hace llegar al destinatario. Cuando la cinta está desenrollada puede verse un conjunto de caracteres que

carece de sentido. Es un ejemplo de sistema criptográfico de transposición. Para que el receptor pueda leer el mensaje adecuadamente debe contar con un bastón que tenga las mismas dimensiones que aquel con el que cifró el mensaje. De esta manera, enrollando de nuevo la cinta y leyendo el texto longitudinalmente es posible recuperar el mensaje original.

Puede observarse por tanto que en este sistema la clave corresponde al diámetro del bastón. A pesar de que hoy en día puede resultar un sistema rudimentario, se considera un sistema ingenioso si se tiene en cuenta que estamos hablando del S. V a.C.

Como curiosidad puede mencionarse que es de aquí de donde viene la entrega del bastón de mando a los alcaldes en los actos de posesión de cargos o también su uso en el entorno judicial o militar, representado el hecho de que la persona que tiene el bastón tiene el poder: la clave para descifrar la información.

1.4.2. Cifrado César

Prosiguiendo hacia delante con la evolución de la criptografía es reseñable el sistema que se utilizaba en tiempos de los romanos. El emperador Julio César utilizaba un cifrado que lleva su nombre. Se trata de un criptosistema monoalfabético que se remonta al S. I a.C.

1.4.3.Funcionamiento

Su funcionamiento es sencillo, consiste en hacer corresponder las letras del alfabeto por las correspondientes cuando dicho alfabeto se ha desplazado una serie de posiciones. Típicamente se hace corresponder la letra A con la D y así sucesivamente, como puede observarse en la siguiente tabla. El desplazamiento se realiza como si se tratara de un buffer circular, es decir, cuando se llega al final (X) se continúa por el principio (A) en la correspondencia.

A	B	C	D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	V	X
D	E	F	G	H	I	K	L	M	N	O	P	Q	R	S	T	V	X	A	B	C

En este sistema la clave se corresponde con el número de posiciones que se ha desplazado el alfabeto para hacer la correspondencia (en este caso 3 ya que se hace corresponder A con D). Esta es la correspondencia típica, pero pueden hacerse otras variaciones.

A continuación, se muestra un ejemplo de este criptosistema.

Mensaje	**V**	**I**	**N**	**I**		**V**	**I**	**D**	**I**		**V**	**I**	**N**	**C**	**I**
Clave	**D**	**D**	**D**	**D**		**D**	**D**	**D**	**D**		**D**	**D**	**D**	**D**	**D**
Criptograma	**B**	**M**	**Q**	**M**		**B**	**M**	**G**	**M**		**B**	**M**	**Q**	**F**	**M**

Dado el mensaje que se quiere cifrar, en este caso “Vini Vidi Vinci”, se escribe la clave de cifrado (en este caso D) bajo cada uno de los caracteres del texto original. Para cifrar el mensaje debe sustituirse cada uno de los caracteres del mensaje por el correspondiente atendiendo a la tabla anterior. De esta manera la letra V se sustituye por la B, la I por la M, la N por la Q, etc. Como puede observarse se trata de un cifrado por sustitución.

Para descrifrar el texto, teniendo conocimiento de la clave, debe realizarse la sustitución contraria, sustituyendo los caracteres del criptograma por los correspondientes en la tabla para obtener el texto claro.

Nótese que al tratarse de un sistema monoalfabético cada aparición de un carácter siempre será sustituido por el mismo carácter. Esto puede observarse en el ejemplo con el carácter I, que siempre es sustituido por el carácter M.

1.4.4.Seguridad y criptoanálisis

Analizando la seguridad de este criptosistema puede verse que en el hecho de que un carácter siempre sea sustituido por el mismo es donde radica una de sus debilidades. Además de que la longitud de la clave es uno y se ésta se repite constantemente, se suma el hecho de que la estadística del texto claro se refleja en el texto cifrado, algo que es indeseable para la seguridad de cualquier criptosistema. Esto hace a este sistema vulnerable a un análisis de frecuencias.

El criptoanálisis de frecuencias se apoya en la frecuencia con la que aparece cada carácter en un determinado idioma, lo cual permite inferir información del texto claro a partir del criptograma.

En caso de que no se conociera el idioma en el que está escrito el texto también podría aplicarse un ataque de fuerza bruta, en el que se prueben sistemáticamente todas las posibles correspondencias hasta obtener un texto claro que tenga sentido.

1.4.5. Edad Media

Durante la Edad Media europea el uso de la criptografía fue muy escaso, casi exclusivamente monjes, raramente monarcas o eclesiásticos. En ocasiones era confundida con esoterismo, magia, ocultismo, etc. y a menudo se menudo empleada para encubrir escritos de este tipo. En general se empleaban sistemas muy triviales, como frases escritas al revés, sustitución de letras por la siguiente en el alfabeto, etc.

Con respecto a los orígenes del criptoanálisis, se cree que fue el análisis textual del Corán, lo que llevó a la invención de la técnica del análisis de frecuencias para romper los cifrados por sustitución monoalfabéticos, en algún momento alrededor del año 1000.

Durante la Edad Media se escribieron varios de libros sobre criptografía y criptoanálisis y se sabe que incluso los Templarios cifraban sus cartas utilizando un método propio.

1.4.6. Renacimiento

Fue en el Renacimiento cuando se produjo la expansión de uso del cifrado. Hubo una crisis de los métodos de sustitución y se generalizó el conocimiento de las frecuencias de aparición de las letras y sílabas.

En esta época surgieron varios de métodos alternativos:

- Silabarios.
- Nomenclatores.
- Nuevos métodos de sustitución.

Veamos a continuación cada uno de ellos.

Los **silabarios** realizan sustitución por sílabas. Consisten en un catálogo de sílabas en el que cada una aparece asociada al símbolo (números, voces, u otras sílabas) que la sustituye en un texto cifrado.

En la figura se puede ver un ejemplo de silabario.

El **nomenclátor** es un catálogo de nombres en el que cada uno aparece asociado a la palabra que le sustituye en un texto cifrado.

Para resolver las vulnerabilidades mencionadas anteriormente los sistemas criptográficos se fueron complicando con el paso del tiempo. Como nuevos métodos de sustitución surgieron la sustitución homofónica y la polialfabética.

Las palabras **homófonas** son aquellas que se pronuncian igual pero se corresponden a diferentes palabras. Siguiendo esta idea, una letra ya no era sustituida por otra letra invariablemente, si no que podía corresponderse a varias diferentes. Esto complica el descifrado del texto y solventa algunas de las desventajas del cifrado monoalfabético.

Se puede ver con el siguiente ejemplo:

Teniendo

C = {06, 29, 48, 63}
I = {04, 13, 19, 71}
N = {22, 42, 56, 67, 76, 84}
O = {05, 25, 35, 49, 81, 89, 94, 96}
S = {08, 12, 21, 53, 60, 74, 83}
T = {01, 61, 77, 92}
U = {17, 26, 98}

La palabra “sustitución” quedaría de esta manera:

S	U	S	T	I	T	U	C	I	O	N
21	98	74	61	04	01	26	29	13	81	76

Como se puede ver la letra “S” no es sustituida siempre por el mismo número, si no que en el primer caso es sustituido por 21 y en el segundo por 74.

Como ejemplo de sustitución **polialfabética** se puede mencionar el cifrado en disco inventado por Leone Battista Alberti (1404-1472). Un ejemplo de este artilugio se puede ver en la figura.

Consiste en dos discos concéntricos unidos en el centro. El disco exterior es fijo y contiene las letras del alfabeto latín convencional ordenado además de las cifras

1, 2, 3 y 4. El disco interior es móvil y contiene otro alfabeto. De esta manera, girando el disco interno se puede hacer corresponder el alfabeto del disco externo con tantos alfabetos como contenga el disco interno, resultando en un sistema polialfabético. De esta forma se dificulta el criptoanálisis del texto.

Veamos un ejemplo de sustitución polialfabética que sigue la idea del **disco de Alberti**. Con letras mayúsculas se representa el alfabeto del disco externo y con minúsculas las del interno. Se va a escoger un alfabeto para cifrar las letras en posiciones pares y otra para las impares.

Correspondencia para posiciones impares:

A	B	C	D	E	F	G	H	I	J
a	e	i	m	p	t	x	b	f	j

K	L	M	N	Ñ	O	P	Q	R	S
n	q	u	y	c	g	k	ñ	r	v

T	U	V	W	X	Y	Z
z	d	h	l	o	s	w

Correspondencia para posiciones pares:

A	B	C	D	E	F	G	H	I	J
n	r	w	b	g	l	p	u	z	e

K	L	M	N	Ñ	O	P	Q	R	S
J	ñ	s	x	c	h	m	q	v	a

T	U	V	W	X	Y	Z
f	k	o	t	y	d	i

El texto “sustitución” quedaría de esta manera:

S	U	S	T	I	T	U	C	I	O	N
v	k	v	f	f	f	d	w	f	h	y

1.4.7.Cifrado Vigenère

En 1586 el francés **Vigenère** propuso un cifrado polialfabético bautizado con su nombre.

La particularidad de los sistemas polialfabéticos es que cada carácter del texto a cifrar no se sustituye siempre por el mismo carácter en el texto cifrado, es decir, es un sistema en el que hay implicados varios alfabetos y dependiendo de ciertas circunstancias se aplica uno u otro. Esto complica a un atacante el criptoanálisis del texto.

1.4.7.1. FUNCIONAMIENTO

Expliquemos este cifrado con un ejemplo.

Se quiere cifrar el mensaje “Paris vaut bien une messe”. Por curiosidad recordaremos que esta frase fue una cita de Enrique IV (1553-1610) de Navarra que quiso ser rey de Francia, pero topó con el impedimento de que él no era católico. Entonces, pronunció esta célebre frase indicando con ello que estaba dispuesto a convertirse al catolicismo para que le dejasen acceder al trono francés.

En este caso la clave de cifrado elegida es: *L, O, U, P*, la cual se repite tantas veces como sea necesario hasta cubrir todo el texto original.

Mensaje	P A R I S	V A U T	B I E N	U N E	M E S S E
Clave	L O U P L	O U P L	O U P L	O U P	L O U P L
Criptograma	A O L X D	J U J E	P C T Y	I H T	X S M H P

Como puede verse en este caso un mismo carácter en el mensaje no se corresponde con el mismo carácter en el criptograma. Para saber a qué carácter se corresponde cada sustitución se utiliza la siguiente tabla:

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

Si nos fijamos atentamente se puede ver que en realidad cada una de las filas se corresponde con un cifrado César, en el que la correspondencia se hace cada vez con una letra del alfabeto.

Para saber la correspondencia del carácter del mensaje con el criptograma la tabla se utiliza de esta manera: en el ejemplo el primer carácter *P* se busca en las columnas y el carácter *L* de la clave se busca en las filas. El carácter correspondiente en la tabla es *A*, que representa el carácter correspondiente en el criptograma. En el caso de *A* con la clave *O* se obtiene *O*. Se procedería de esta manera con todo el texto.

Como se puede observar, el mismo carácter en el criptograma no se corresponde siempre con el mismo carácter en el criptograma como sucedía con el cifrado monoalfabético. Por ejemplo, la *A* de la palabra *VAUT* se corresponde con *U* cuando anteriormente se correspondía con *O*. Sucede lo mismo con el carácter *E* u otros que se quieran analizar. Esta es una de las ventajas del cifrado polialfabético, sin embargo, esto no ha impedido que este criptosistema sea criptoanalizado.

Por supuesto, cuanto mayor sea la longitud de la clave más seguro será el sistema. El cifrado de Vigenère puede atacarse haciendo análisis de frecuencias a los diferentes cifrados César.

Cabe preguntarse, ¿cuál es el número de posibles claves en función de su longitud?

Considerando un alfabeto de 26 letras se tiene:

Longitud de la clave	Número de posibles claves
1	26
2	26*26= 676
3	676*26 = 17 576

Y así se seguiría sucesivamente.

1.4.7.2. SEGURIDAD Y CRIPTOANÁLISIS

Hay básicamente dos métodos para atacar el cifrado de Vigenère. El primero es un ataque por diccionario. Se trata de un ataque de fuerza bruta en el que se prueba exhaustivamente cada palabra de diccionario como clave del criptosistema. Dejando a un lado el tiempo y los recursos computacionales necesarios para probar todas las posibilidades, que en un ordenador a día de hoy no debería ser demasiado, este método tiene una clara limitación, y es que solo tendrá éxito en caso de que la clave sea una palabra que se encuentre en el diccionario. Si se utiliza cualquier otra clave no será posible obtener el texto claro.

Es por ello que matemáticos, criptoanalistas y científicos se enfrentaron al problema del criptoanálisis del cifrado de Vigenère, el cual era conocido como la

cifra indescifrable. El objetivo era poder criptoanalizar texto cifrado incluso cuando la clave no es palabra contenida en el diccionario, incluso si se trata de un texto aleatorio.

En 1854 Charles Babbage logró criptoanalizar este sistema, sin embargo no publicó sus resultados. Fue en 1863 cuando Kasiski publicó estos resultados en el libro "La escritura secreta y el arte de descifrar".

Este método se basa en la redundancia del lenguaje. Si un texto se lee a saltos entre las letras, las estadísticas de frecuencia de aparición de las letras en esos subtextos serán similares a las del texto original. Es decir, las frecuencias de cada letra se conservan en los textos separados un determinado número de espacios (*X*). Para ello se debe contar con una cantidad de texto suficiente para que estas estadísticas den resultados fiables.

¿Cuánto texto es necesario para que se manifieste la redundancia del lenguaje? En un alfabeto de 27 letras, con 100 letras sería suficiente. Con 3 veces el tamaño del alfabeto las estadísticas ya empiezan a ser fiables.

En general, si la clave es de longitud *X*, con un texto de 100*X* caracteres se tendrían buenas probabilidades de que el ataque de Kasiski tuviese éxito.

1.4.8. Una curiosa historia

Para ilustrar la relevancia que ha tenido la criptografía en la historia, así como las consecuencias que pueden llegar a tener determinados fallos de seguridad en sistemas criptográficos, se relata a continuación una historia real. La historia tiene como protagonistas a María Estuardo, Reina de Escocia (1542-1587) y su prima, Isabel Tudor, Reina de Inglaterra (1533-1603).

María estuvo encerrada e incomunicada durante 18 años puesto que los católicos en Inglaterra no la consideraban la reina legítima, lo contrario de lo que pensaban de María. Un día recibió unas cartas a través de Gilbert Gifford, que consiguió colarlas en la prisión con técnicas esteganográficas. Además, consiguió que María pudiera enviar correspondencia sin que fuera delatada. Se puede decir que Gifford era un punto intermedio entre los mensajes de/a María (MITM).

A su vez, un grupo de nobles católicos liderados por Anthony Babington planeaba derrocar a Isabel I y liberar a María de Escocia. Querían comunicar a María su plan para que diera su aprobación.

Gifford se puso en contacto con Babington y lo utilizó para hacer llegar a María la carta con su plan. La carta estaba escrita utilizando un nomenclátor para mantener el secreto. Gifford entregó la carta a María, pero también a gente cercana a la reina Isabel I, los cuales descubrieron el plan pero decidieron esperar a la contestación de María.

María respondió usando el nomenclátor haciendo algunos ajustes sobre el plan, lo cual supuso la firma de su sentencia de muerte.

Antes de entregar la carta a Babington, añadieron la siguiente postadata:

"Me alegraría conocer los nombres y las cualidades de los seis caballeros que llevarán a cabo el plan; porque puede que, conociendo a los participantes, yo pueda daros algún consejo necesario para seguirlo en eso, así como de vez en cuando los particulares de cómo proceder: y en cuanto podáis, con el mismo propósito, quienes conocen ya, y en qué medida los detalles de esta cuestión". Babington no desconfío de la carta y proporcionó la información, finalizando la historia con la muerte de todos los implicados en la conspiración, incluida María, que fue decapitada en 1587.

Esta historia refleja la importancia de no utilizar sistemas de cifrado débiles, pues confiar en su cifrado puede llegar a ser más peligroso que no cifrar el mensaje.

1.4.9. Cifrado Playfair

La criptografía ha estado muy ligada al ámbito militar. En la primera guerra mundial se utilizó en los frentes de batalla.

En 1854 se inventó el cifrado Playfair, ideado por Charles Wheatstone y que toma este nombre por el varón que lo promovió. Se trata de la primera cifra de sustitución digráfica.

Utiliza una clave formada por una matriz de cifrado 5x5.

Contiene una serie de reglas para la sustitución. Para cifrar 2 caracteres m1, m2:

• Si **m1** y **m2** se encuentran en la misma fila, escoger **c1** y **c2** situados a su derecha (circularmente).

• Si **m1** y **m2** se encuentran en la misma columna, escoger **c1** y **c2** situados debajo (circularmente).

• Si **m1** y **m2** se encuentran en distintas filas y columnas, escoger **c1** y **c2** situados en la diagonal opuesta

(m_{ij} m_{kl}-> c_{il} c_{kj}).

• Si **m1** = **m2**, insertar carácter sin significado entre **m1** y **m2** para evitar su repetición, y después aplicar las reglas 1-3.

• Si el número de letras es impar, añadir una sin significado al final del texto.

Veamos su funcionamiento con un ejemplo:

Dado el mensaje: "sustitución poligráfica", se agrupan las letras de dos en dos:

SU ST IT UC IO NP OL IG RA FI CA. Siendo ésta la matriz:

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

el mensaje cifrado queda así: FE GK NK PA LS CB HO NS UL SA AN.

Para ello se han sustituido "SU" por "FE", ya que se trata de letras que se encuentran en diferentes filas y columnas, y por tanto se sustituyen por las que se encuentran en la diagonal opuesta. En el caso del último par "CA", al estar en la misma fila, se sustituye por "AN" que son las que se encuentran a su derecha en la matriz.

Este sistema fue vulnerado por Mauborgne (1914).

1.4.10. La II Guerra Mundial

Siguiendo con la trayectoria bélica, la II Guerra Mundial fue otro de los escenarios donde la criptografía fue protagonista. Se construyeron varias máquinas de cifrado, como las máquinas Enigma y Lorentz en Alemania. Hubo otras, como Púrpura y Jade en Japón.

Por su popularidad e importancia en la historia se va a ahondar más en el cifrado de la **máquina Enigma**. Esta máquina se utilizó durante la Segunda Guerra Mundial para cifrar y descifrar los mensajes que se intercambiaban durante la guerra.

En la foto se muestra un ejemplar de **máquina Enigma**. Como puede verse, es parecida a una máquina de escribir. Está compuesta por un tablero de clavijas, el teclado, un panel luminoso en el que aparece iluminada la letra a la que se corresponde cifrada la tecla pulsada, además de los modificadores y rotores.

Es una máquina que realiza varias operaciones para el cifrado de cada carácter, lo que hace que el número de combinaciones posibles sea ciertamente elevado. Además, se utilizaba una clave de cifrado que cambiaba cada 8 horas, lo que complicaba aún más su criptoanálisis.

Alan Turing, considerado uno de los padres de la computación, junto a otros expertos, fueron protagonistas en el proceso de criptoanálisis de la máquina Enigma para descifrar los códigos nazis. Fue en Bletchley Park donde construyeron la llamada "**bomba de Turing**" y junto a su potencia y algunos fallos en el uso de Enigma fue posible descifrar los mensajes de los alemanes en la Segunda Guerra Mundial. Se estima que este hecho fue capaz de acortar en 2 o 3 años la duración de la guerra.

Otra de las técnicas usadas durante la Segunda Guerra Mundial fue la utilización de lenguas poco extendidas para ocultar de esta forma los mensajes secretos. En particular se usó el idioma de los indios navajos. Se estima que lo hablaban únicamente unos 4000 indios, por lo que era una lengua bastante desconocida y difícil de entender por el resto del mundo. El código estaba hecho a de palabras o frases en idioma navajo, que luego era traducido al inglés.

Éstos son algunos de los aspectos más destacables de la historia de la criptografía, aunque obviamente hay muchos más que no se han reflejado aquí en pro de no extenderse demasiado. El alumno interesado puede investigar más por su cuenta.

Todos estos hechos forman parte de lo que se llama la criptografía precientífica, la cual tenía en cierta medida una componente artística. Es a partir de que Shannon en 1949 propusiera sus postulados cuando se empieza a hablar de la criptografía científica. Estos aspectos se estudiarán con más detalle en el resto de esta asignatura.

1.5. Seguridad criptográfica

En relación a la seguridad criptográfica se puede distinguir entre los siguientes conceptos:

- **Incondicional** (teórica). El sistema es seguro frente a atacantes con tiempo y recursos computacionales ilimitados. Actualmente el único sistema conocido que cumple estas características es el cifrado de Vernam.
- **Computacional** (práctica). Es seguro frente a atacantes con tiempo y recursos computacionales limitados. Un ejemplo de este tipo de sistemas es RSA.
- **Seguridad probable**. No se puede demostrar su seguridad, pero el sistema no ha sido violado. Un ejemplo es el algoritmo DES.
- **Seguridad condicional**. Los demás sistemas. Por ejemplo, los algoritmos clásicos, que eran seguros mientras que un atacante no tenía medios para romperlos.

Una de las diferencias entre la criptografía clásica y la moderna es que la primera ofrecía sistemas con seguridad probable mientas que, a día de hoy, con los avances en el criptoanálisis y la capacidad cada vez mayor de los sistemas

computacionales es necesario que la seguridad sea matemáticamente demostrable.

1.6. Criptoanálisis

Como se comentó el criptoanálisis persigue romper el cifrado y obtener ilegítimamente el mensaje original a partir del criptograma.

Antiguamente, para preservar la seguridad de los mensajes intercambiados, el funcionamiento del procedimiento criptográfico se escondía para evitar que los atacantes conocieran el método y pudieran emplearlo para leer los mensajes. La desventaja que tiene este método es que, si el algoritmo de cifrado fuese revelado, filtrado o se utilizase ingeniería inversa, un atacante podría tomar ventaja de esto y comprometer el mensaje original.

Para evitar esto en el S XIX se enunció el **principio de Kerckhoffs**, que dice lo siguiente:

"El atacante tiene pleno conocimiento del método de cifrado con excepción de la clave".

Esto da un vuelco al concepto anterior. De hecho, a día de hoy se garantiza la seguridad de los algoritmos de cifrado al hacerlos públicos y retar a que se rompan. De este modo la seguridad reside únicamente en la clave de cifrado y no en el ocultismo del modo de funcionamiento del algoritmo.

Los ataques pueden clasificarse en función a varios criterios. La primera clasificación diferencia entre los ataques pasivos y los activos.

- Los **ataques pasivos** son aquellos en los que el atacante monitoriza el canal de comunicación (ataque a la confidencialidad) pero no interviene sobre la información.
- **Ataques activos**. A diferencia de los anteriores, en este caso el atacante intenta modificar la información intercambiada, añadiendo, borrando o alterando los mensajes enviados (esto supone un ataque a la confidencialidad, integridad y autenticidad).

Otra clasificación es la siguiente:

- **Ataque sobre texto cifrado únicamente**. En este ataque pasivo el atacante solo tiene acceso al criptograma. Tiene un conocimiento mínimo del mensaje original (como el idioma). Un buen método de cifrado no debería ser vulnerable a este ataque.
- **Ataque sobre texto claro conocido**. En este caso el atacante conoce una porción del texto claro, así como su correspondiente texto cifrado. Partiendo de esta información el atacante tiene como objetivo el deducir la clave o descifrar una nueva fracción del texto claro.
- **Ataque sobre texto claro elegido**. Se trata de un ataque activo en el que el atacante puede descifrar un texto claro de su elección. Esto podría suceder cuando el atacante tiene acceso al dispositivo de cifrado o puede enviar textos claros de su elección para cifrarlos y que posteriormente el texto cifrado se envíe a un tercero.
- **Ataque adaptativo**. Este ataque es una variante del anterior. En este caso el atacante puede obtener nuevo texto cifrado en función de otros criptogramas obtenidos anteriormente. Todos los textos cifrados corresponden a textos claros de su propia elección. Aunque este escenario no es muy realista tiene interés teórico.

La eficacia de un ataque criptoanalítico se mide por varios factores:

- Cantidad de pares texto claro y texto cifrado necesarios
- Tiempo para criptoanálisis
- Probabilidad de éxito del ataque.

1.6.1. Ejemplo de criptoanálisis

Un ejemplo de criptoanálisis se encuentra en el libro “El escarabajo de oro”, escrito por Edgar Alan Poe en 1843. Este libro explica detalladamente cómo se puede romper un procedimiento de cifrado mediante técnicas estadísticas. La obra relata una historia en la que se debe descifrar un mensaje secreto para encontrar la localización de un tesoro escondido.

El criptograma que aparece en el libro es el siguiente:

53‡‡†305))6*;4826)4‡.)4‡);806*;48†8¶

60))85;1‡(;:‡*8†83(88)5*†;46(;88*96*?;

8)*‡(;485);5*†2:*‡(;4956*2(5*-4)8¶8*

;4069285);)6†8)4‡‡;1(‡9;48081;8:8‡1;48

†85;4)485†528806*81(‡9;48;(88;4(‡?34

;48)4‡;161;:188;‡?;

Vamos a ver de qué manera es posible criptoanalizar este criptograma.

Realizando un análisis de frecuencias se sabe que la letra más usada en la lengua inglesa es la letra “e”.

Analizando el criptograma se puede ver que el carácter más repetido es el “8”, por tanto, se hace la asociación e -> 8. De esta forma se sustituyen todas las apariciones de 8 en el criptograma por la letra “e”. El texto quedaría como sigue:

53‡‡†305))6*;4e26)4‡.)4‡);e06*;4e†e¶

60))e5;1‡(;:‡*e†e3(ee)5*†;46(;ee*96*?;

e)*‡(;4e5);5*†2:*‡(;4956*2(5*-4)e¶e*

;40692e5);)6†e)4‡‡;1(‡9;4e0e1;e:e‡1;4e

†e5;4)4e5†52ee06*e1(‡9;4e;(ee;4(‡?34

;4e)4‡;161;:1ee;‡?;

Para continuar se tiene en cuenta que la palabra “the” es la más habitual en el idioma inglés. Por tanto, se debe buscar en el criptograma el patrón **8, siendo * cualquier carácter. Se ve que en el criptograma aparece 7 veces repetido el patrón “;48”, de lo cual se deduce la siguiente asociación

t -> ;
h -> 4
e -> 8

En el criptograma se puede ver ";(88;4". Tenemos todos los caracteres correspondientes excepto "(". Como la "h" la asociamos a "4", se deduce que el "(" no es "h". Entonces se reduce a ";(88". Esto encaja con la palabra "tree", lo cual relaciona

r -> (

Sustituyendo se obtiene: "the tree thr***h the". De donde se deduce la palabra "trough" que da lugar a las asociaciones

O -> ‡

u -> ?

g -> 3

Buscando en el texto se encuentra 83(88)

egree. Asociando la palabra degree se obtiene una nueva letra

d -> +

Luego se ve la combinación 46 (; 88. Sustituyendo se obtiene th * rtee*, lo cual sugiere la palabra "thirteen" y que permite obtener más asociaciones

i -> 6

n -> *

De "5good" se deduce

a -> 5

Sustituyendo los símbolos en el criptograma se obtiene el texto original:

"A good glass in the bishop's hostel in the devil's seat forty-one degrees and thirteen minutes northeast and by north main branch seventh limb east side shoot from the left eye of the death's-head a bee line from the tree through the shot fifty feet out."

Lo cual nos lleva a la localización del tesoro ☺

1.7. Condiciones de secreto perfecto

En 1949 **Shannon** enunció las condiciones de secreto perfecto, las cuales reflejan en forma matemática la seguridad de un sistema criptográfico. Estas condiciones

se basan en dos hipótesis fundamentales

1. La clave se utilizará solamente una vez.

2. El atacante sólo tiene acceso al criptograma.

Tomando esto como base, se dice que un sistema criptográfico cumple las condiciones de secreto perfecto si el texto claro X es estadísticamente independiente del criptograma Y, para todos los posibles textos claros y todos los posibles criptogramas.

Matemáticamente se expresa como

$$P(X = x \mid Y = y) = P(X = x)$$

Esto refleja que el valor tomado por la variable X es el mismo con o sin el conocimiento de la variable Y. Dicho de otra manera, el criptograma no aporta información sobre el texto claro a un criptoanalista, independientemente de los recursos y tiempo de los que disponga.

Shannon, conocido como padre de la teoría de la información, introdujo el concepto de entropía. Ésta mide la incertidumbre de una fuente de información y es un concepto bastante usado en criptografía. Basándose en la entropía, Shannon determinó cuál era la **longitud mínima de la clave** para que se pudieran dar las condiciones de secreto perfecto. El resultado es que la longitud de la clave tiene que ser al menos tan larga como la longitud del texto claro. Esto es, $K >= M$, siendo K la longitud de la clave y M la del mensaje original.

1.8. Cifradores aditivos

Los cifradores aditivos módulo L verifican las condiciones de cifrado perfecto.

Dado un alfabeto {0, 1, ..., L-1} para el texto claro y criptograma, sea X el texto claro, Y el texto cifrado, Z la clave, M la longitud del texto claro, N la longitud del criptograma y K la longitud de la clave,

Debe darse que $M=K=N$.

La clave Z debe elegirse de forma aleatoria.

La transformación de cifrado ser realiza aplicando una operación XOR entre el cada elemento del texto claro y la clave.

A cada texto claro X le puede corresponder, con igual probabilidad, cualquiera de los L^M posibles criptogramas. Por tanto, la información aportada por el criptograma sobre el texto claro es nula, esto es, X e Y son estadísticamente independientes.

1.9. Aplicaciones de la criptografía

A pesar de que la Criptografía pueda parecer estar alejada de nosotros en realidad está presente en nuestro día a día y en muchas más acciones cotidianas de las que se pueda pensar en un primer momento.

La criptografía está presente en acciones tan cotidianas como éstas: utilizar un cajero automático, realizar una llamada con nuestro teléfono móvil, usar aplicaciones de mensajería tipo Telegram o Whatsapp, intercambiar ficheros o fotos entre dispositivos, ver la TV de pago, utilizar la administración electrónica, almacenar información confidencial, identificarnos con el DNI-e/ pasaporte-e, votar electrónicamente, usar marcas de agua, etc.

La criptografía también tiene aplicaciones en protocolos de seguridad, como PGP, SSH (HTTPS, HSTS, TOTP), criptomonedas (ej. Bitcoin), redes anónimas (como Tor), y un largo etcétera.