Protección y Seguridad de la Información

Laura M. Castro Souto

Segundo Cuatrimestre
Curso 2000/2001

Índice general

Capítulo 1

Introducción y Fundamentos

1.1. Conceptos Generales: Fundamentos básicos de la Seguridad de la Información

Los objetivos fundamentales de la **Seguridad de la Información** son dos:

- Mantener el secreto de cara a los accesos a la información.
- Mantener la autenticidad evitando modificaciones no autorizadas de la información.

Normalmente un flujo de información une una fuente y un destino:

Partiendo de este sencillo esquema, podemos clasificar ó hablar de 4 categorías de *ataques*:

- Interrupción.
- Intercepción.
- Modificación.
- Fabricación.

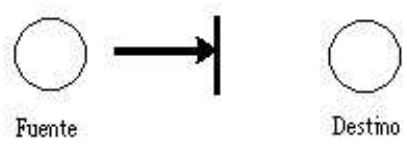
Veremos en qué consiste cada uno de ellos.

1.1.1. Tipos de ataques. Defensa.

Interrupción

Las consecuencias de este tipo de ataque son típicamente la **destrucción** y/o **inutilización** de la información.

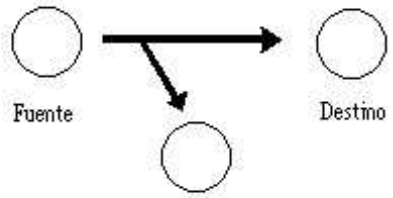
Es un ataque, pues, sobre la *disponibilidad*.



Ejemplos: destrucción de un elemento hardware (ataques físicos), corte de una línea de comunicaciones (deliberada o accidentalmente), borrado de ficheros, registros, bases de datos, programas...

En caso de un fallo de este tipo, es necesario detectarlo convenientemente, evaluarlo y sobre todo actuar rápidamente, momento en el cual intervendrán diversos factores, entre ellos el económico.

Intercepción



La **intercepción** consiste en la **participación sin autorización** realizada por personas, computadoras o en general cualquier tipo de entidad, en la comunicación entre la fuente y el destino de la información.

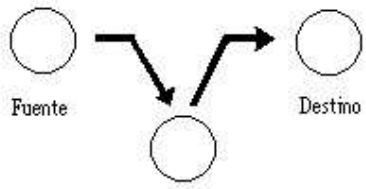
Es el ataque más difícil de detectar, ya que todo funciona bien, no como en el caso anterior; aquí se está atentando contra la *confidencialidad*. Ejemplos: *sniffers*, copia de software, etc.

El tipo de red usado está íntimamente relacionado con la seguridad en este sentido. En una *red de medio compartido* (Ethernet), todos los ordenadores lanzan sus paquetes al medio físico, de modo que un sniffer conectado a la misma podría verlos sin restricción alguna. Existen técnicas software de cifrado (SSH, SSL, PGP,...) para estos casos, aunque actualmente se está evolucionando hacia un método más eficiente como es el cifrado a nivel IP (un único cifrado para todo tipo de aplicaciones, con IPv6). Por contra, en una *red de medio conmutado* ó *switcheado* (switching Ethernet), se establecen conexiones a nivel de enlace, de modo que ya no tenemos ese problema.

Modificación

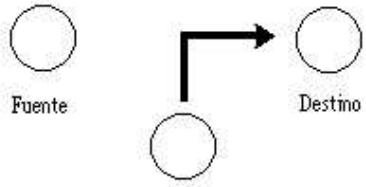
En un ataque de **modificación** se captura la información, que es tratada y posteriormente reenviada al destino.

También se da, como en el caso anterior, una participación sin autorización: se accede a la información pero además se modifica, se altera.



Es un ataque contra la *integridad* de los datos, por lo que resulta muy peligroso. Ejemplos: cambios en valores de BD, programas, modificación de mensajes, troyanos...

Fabricación



En ocasiones como ésta, el intruso o elemento subversivo intenta hacerse pasar por la fuente, siendo, pues, un ataque sobre la *autenticidad*. Se introducen en el sistema objetos o entidades fabricadas. Ejemplos: introducción de mensajes en una red, inclusión de campos o tablas en una BD, inclusión de virus[1],...

Otra clasificación para los tipos de ataques puede ser:

- Activos.
- Pasivos.

Ataques Activos

Los ataques activos suelen modificar la información, los datos o los mensajes. Pueden consistir, por ejemplo, en el cambio de la identidad de un emisor/receptor, la manipulación de datos, denegación de servicios, encaminamiento incorrecto o incluso la **repetición**:

[1]Este ataque concreto ha ido evolucionando: en principio se transmitía por medio de los dispositivos de almacenamiento, y ahora tiene en Internet, y sobre todo en el correo electrónico, su principal aliado.

Según la *Teoría de la Información*:

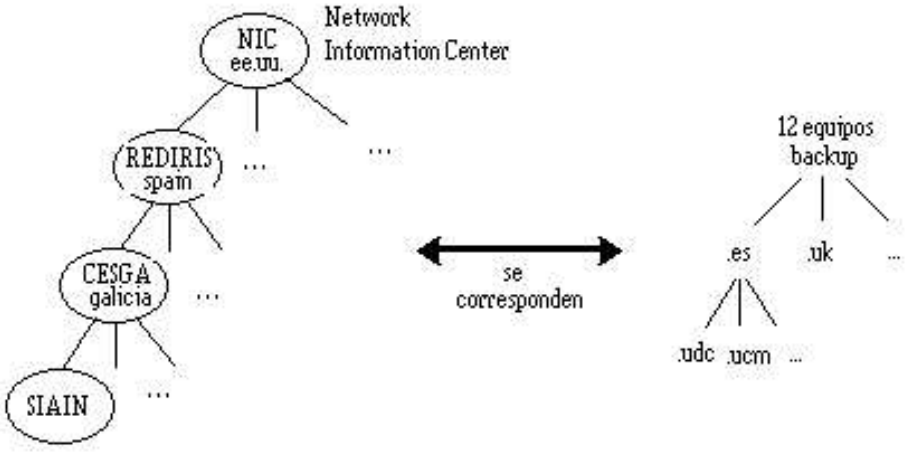
> "Se adquiere información cuando se conoce algo que antes no se sabía. Es el cambio que se produce entre el desconocimiento o incertidumbre de un hecho y el conocimiento o la certidumbre del mismo".

Ataques Pasivos

Pueden consistir, básicamente en:

a) *Observación de mensajes* (simple acceso a la información), o bien

b) *Análisis de tráfico* (no se accede a la información, que suele ir cifrada, sino que se roba información sobre el tráfico: tipo de frecuencias de envío, identificación de usuarios, y en general, características del intercambio entre sistemas que pueden usarse después en acciones como el reemplazo de IPs en cachés de servidores DNS).



1.2. Antecedentes Históricos y Evolución

La información siempre ha significado dinero y/o poder. Por ello, la demanda de la seguridad de la información ha provenido históricamente de aquéllos que han tenido información que proteger o bien intereses en conseguirla, como han sido el clero, los militares, banqueros, gobiernos y sistemas políticos.

Haremos a continuación una revisión rápida de los acontecimientos y nombres más destacados en este ámbito hasta la actualidad.

Según el Diccionario de la Real Academia, la palabra **Criptografía** proviene del griego *kriptos*, que significa oculto, y *graphos*, que significa escritura, y su definición es: *"Arte de escribir con clave secreta o de un modo enigmático"*. Es la inclusión de la reseña "clave secreta" la que marca la diferencia entre *codificado* y *cifrado*. Obviamente la Criptografía hace años que dejó de ser un arte para convertirse en una técnica, o más bien un conglomerado de técnicas, que tratan sobre la protección —ocultamiento frente a observadores no autorizados— de la información. Entre las disciplinas que engloba cabe destacar la Teoría de la Información, la Teoría de Números —o Matemática Discreta, que estudia las propiedades de los números enteros—, la Estadística y la Complejidad Algorítmica ó Teoría de la Complejidad Computacional. Hoy en día se considera una ciencia aplicada en toda regla, debido a dicha relación con otras ciencias.

Otras definiciones que se han dado han sido, por ejemplo: *"Escritura secreta relacionada mediante una clave, indispensable para descifrarla"*.

Existen dos documentos fundamentales, uno escrito por Claude Shannon en 1948 (*"A Mathematical Theory of Communication"*), en el que se sientan las bases de la Teoría de la Información, y que junto con otro artículo posterior del mismo autor (*"Teoría de las comunicaciones secretas"*[2], 1949) sirvió de base para la Criptografía moderna. El segundo trabajo fundamental, publicado por Whitfield Diffie y Martin Hellman en 1976, se titulaba *"New directions in Cryptography"*, e introducía el concepto de Criptografía de Clave Pública, abriendo enormemente el abanico de aplicación de esta disciplina (a él debemos, pues, los orígenes de los *métodos asimétricos*).

Conviene hacer notar que la palabra Criptografía sólo se refiere al uso de códigos, por lo que no engloba a las técnicas que se usan para romper dichos códigos (**Criptoanálisis** —ciencia dedicada a quebrantar el cifrado—). El término **Criptología**, aunque no está recogido aún en el Diccionario, se emplea habitualmente para agrupar estas dos disciplinas.

Sin olvidar las excavaciones en Egipto y los hallazgos que han producido (*Piedra Rosetta...*), que datan de 1900 años a.C., el primer uso de la escritura secreta de que se tiene constancia data del s.V a.C., durante la guerra entre Atenas y Esparta. En aquel caso, el cifrado se basaba únicamente en la alteración del mensaje mediante la inclusión de símbolos innecesarios que desaparecían al enrollar la lista en un rodillo, de manera que se mostraban sólo los símbolos del mensaje, pudiéndose así leer fácilmente.

Otra de las primeras noticias sobre criptografía proviene de la época de los romanos. El cifrado en ese caso consistía en una sustitución de determinados símbolos por otros según una regla fija (método denominado **método César**, en honor a Julio César), un método de **sustitución** no demasiado seguro, pues como todos los de este tipo es fácil de descifrar desde utilizando tablas de frecuencias (ver tabla 1.2 en página 12) hasta por *fuerza bruta* (con cualquier método que explore exhaustivamente el *espacio de claves*, esto es, en este caso, con todos los posibles desplazamientos; para evitar esto, la evolución más natural a lo largo de la historia fue aumentar el espacio de claves de manera que fuese siendo cada vez más difícil —más "caro", computacionalmente hablando— romperlas).

[2]Esta teoría fue aplicada por el NBS —National Bureau of Standards— de EE.UU. para desarrollar el sistema criptográfico DES —Data Encryption Standard—.

K (clave) = 3

A	B	C	D	E	F	G	H	...
↓								
D	E	F	G	H	I	J	K	...

Cuadro 1.1: Ejemplo del *método César*.

Augusto utilizaría otro método en el que el texto usado como clave era el propio texto a cifrar (este método se recuperaría muchos siglos más tarde para usos como el famoso *teléfono rojo* de la Casa Blanca). Claro que esta estrategia no soluciona el problema de las frecuencias, porque la clave también las tiene (la única alternativa es usar un texto *aleatorio* como clave).

Letra	Español	Inglés	Letra	Español	Inglés
A	11,970 %	8,105 %	Ñ	0,074 %	
B	1,000 %	1,477 %	O	9,195 %	7,389 %
C	4,919 %	2,807 %	P	3,445 %	1,913 %
D	5,190 %	4,221 %	Q	0,875 %	0,105 %
E	13,650 %	12,676 %	R	6,696 %	5,641 %
F	0,953 %	2,356 %	S	7,983 %	6,593 %
G	1,093 %	1,773 %	T	4,802 %	9,634 %
H	0,585 %	5,869 %	U	3,996 %	3,036 %
I	6,860 %	7,176 %	V	0,693 %	0,920 %
J	0,272 %	0,088 %	W	0,019 %	2,330 %
K	0,022 %	0,667 %	X	0,183 %	0,208 %
L	5,270 %	3,964 %	Y	0,523 %	1,570 %
M	2,925 %	2,436 %	Z	0,291 %	0,069 %
N	6,690 %	7,036 %			

Cuadro 1.2: Tabla de frecuencias del español y el inglés.

[3]

También los hebreos y los griegos habían experimentado en este campo. Los primeros utilizaban un sistema sin clave, el *atbash*, consistente simplemente en sustituir cada letra del alfabeto con su simétrica (A→Z, B→Y,...). Los segundos (203-130 a.C.) emplearon por primera vez un cifrado también sin clave en el que se sustituían *letras por números*. Usaban representaciones matriciales (algo que siglos más tarde repetirá el método DES) similares a:

	1	2	3	4
1	A	B	C	...
2	E	F	...	
3	...			

[3] Curiosidad: ¿Qué ocurriría si en un idioma todas las letras fuesen equiprobables?

Más tarde, ya en el s.VIII d.C., los árabes también se especializaron en la aplicación de tablas de códigos en las que un carácter del texto cifrado se correspondía con n del texto original. Eran los primeros **sistemas no homófonos**.

Sin embargo, la obra más antigua que se conserva sobre Criptografía es del s.XIV. Se titula *Liber Zifrorum* y su autor, Cicco Simoneta, estudia en ella diversos sistemas basados en simples sustituciones de letras. Durante toda la Edad Media este campo de estudio estará completamente en manos de frailes y monjes (igual que el resto del saber de la época). Se pone especial interés en el estudio de los diferentes alfabetos y códigos. El italiano Gabrieli di Lavinde formula la primera nomenclatura de la materia.

En el siguiente siglo, Alberti (1404-1472) destaca en el campo del Criptoanálisis, siendo considerado por ello el padre de la Criptología. Durante el siglo XVI se generaliza el uso de la Criptografía en los ambientes diplomáticos y en 1586 Blaise de Vigenère publica una obra titulada *Traicté des Chiffres*, donde recoge diferentes métodos utilizados en la época. Otros nombres destacados son Leon de Baptista, que estudiando la "composición" del latín argumenta que un texto en el que haya aproximadamente 300 vocales, ha de tener unas 400 consonantes. Hace también los primeros estudios de frecuencias e incluye el carácter nulo en los alfabetos de cifrado, lo que los complica un poco más.

Figura 1.1: Ruedas de cifrado de Leon Baptista.

Johanes Trithemieos estudió los sistemas de claves distribuidos matricialmente, así como los sistemas **polialfabeto**. Giuvan Batista acuña el término **clave**, que se extenderá univesalmente, y se da cuenta de que no es sino en ella donde reside la base de todo sistema de cifrado.

El siglo XVI es, pues, el siglo de los estudios sistemáticos. Giovani Puerta de Batista es el primero en hacer una clasificación exhaustiva de todos los métodos criptográficos conocidos hasta entonces, y curiosamente los clasifica tal y como lo hacemos hoy en día:

$$\begin{cases} \text{Clásicos} \\ \text{Modernos} \begin{cases} \text{Simétricos} & \text{(o de Clave Privada)} \\ \text{Asimétricos} & \text{(o de Clave Pública)} \end{cases} \end{cases}$$

En sus estudios estuvo a punto de descubrir un algoritmo para desvelar la longitud de las claves utilizadas en sistemas polialfabeto (posibilitando así la descomposición del texto cifrado en n sistemas monoalfabeto, donde n sería el tamaño de dicha clave). Esta tarea sería completada 200 años después por Kasinski. Aconsejó, tal y como hacemos hoy

en día, el uso de claves largas (reseñando que la situación ideal sería tener una clave tan larga como el propio texto a cifrar), el uso de sinónimos en los textos (para evitar que se capturen frecuencias en un texto cuyo tema o contenido se conoce o se intuye) y el cambio frecuente de la clave.

La idea de utilizar el propio texto como clave fue recuperada por Giovano Cardano, autor del primer libro de probabilidad. Sin embargo, también se dio cuenta de que esta estrategia no era óptima, desarrollando a raíz de esto una variante: utiliza una clave (carácter) para cifrar el primer carácter, el obtenido para cifrar el siguiente, y así sucesivamente. Este método sería usado y considerado infalible hasta la Primera Guerra Mundial.

En el s.XIX se utiliza ampliamente un método basado en la reordenación de los símbolos del mensaje, llamado **transposición**, que junto a la sustitución constituye la base de los cifrados clásicos. Es este un siglo de grandes avances en Criptografía, donde destacan varios nombres propios: Charles Babbage (1791), precursor de las computadoras actuales. Diseña en 1823 la Máquina Diferencial 1, y más tarde la Máquina Diferencial 2, en la que incorpora la revolucionaria idea de una memoria y un procesador por separado. Podemos decir así que la Criptografía moderna nace al mismo tiempo que las computadoras. Babbage fue también quien consiguió romper la *cifra de Vigenère*. Thomas Jefferson (más tarde presidente de los Estados Unidos) fue el precursor de muchos métodos usados posteriormente en las Guerras Mundiales. Bajo su mandato se construyó una máquina codificadora mecánica con 36 ruedas dentadas cuyo espacio de claves era del orden de $37 \cdot 10^{39}$. La mayor contribución de Kasinski (1805-1881) fue el método que lleva su nombre, que desvela la longitud de la clave en sistemas polialfabeto. Kerchoffs (1835-1903) fue el autor de un importante tratado sobre criptografía militar. Afirmó, lleno de razón: *"Un sistema ha de mantener ante todo su impenetrabilidad"*, en referencia no sólo a su clave, sino a la manera que tiene de hacer las cosas (algoritmo).

En siglo XIX el medio de comunicación por excelencia, el que alcanzó mayor auge, fue sin duda alguna la radio (utilizada por los militares en sus comunicaciones). Pero las ondas, la transmisión por aire, son también el medio más propicio al *sniffing*, así que es comprensible el renacer que se dio en Criptología. No obstante, el siglo más activo de la historia en este campo iba a ser el siglo XX. Como en muchas otras áreas científicas, el mayor desarrollo de la Criptología tuvo lugar, como venimos dejando entrever, durante las dos guerras mundiales. En este caso se debió a la necesidad de establecer comunicaciones secretas militares y diplomáticas utilizando nuevas tecnologías, como la telegrafía y la radiotecnia.

En sus inicios esta actividad estuvo marcada por el enfrentamiento entre dos potencias: Francia y Alemania. Los franceses crearon dos grupos de trabajo especializados, uno en Criptografía y otro en Criptoanálisis. Como hitos fundamentales, cabe destacar la intercepción y descifrado por parte de los ingleses de los télex alemanes al embajador germano en EE.UU., Tedesco Zimmerman, durante la Primera Guerra Mundial.

Ya durante la Segunda Guerra, en un lugar llamado Bletchley Park, un grupo de científicos entre los que se encontraban Alan Turing y Vonn Neuman, trabajaba en el proyecto ULTRA tratando de descifrar los mensajes enviados por el ejército alemán con el más sofisticado ingenio de codificación ideado hasta entonces: la máquina ENIGMA, obra de Scherbius. Esta máquina (patentada en 1918) contaba con diversos *modificadores*

que giraban $1/n$ de vuelta (donde n era el número de caracteres del alfabeto utilizado) cada vez que se codificaba un carácter (de suerte que cada n caracteres codificados volvían a su posición original). La inclusión de más de uno evitaba la repetición secuencial de una única clave, siendo n^2, n^3,... el número de posiciones posibles en función del número de modificadores (2,3,...). Si añadimos a esto la posibilidad de cambiar el orden de los modificadores, se obtiene un espacio de claves del orden de 10^{16}. La clave de este sistema era, pues, obviamente, dicha posición inicial de los modificadores. Las claves se recogían en un *Libro de códigos* y cada día los alemanes empleaban una distinta en sus transmisiones.

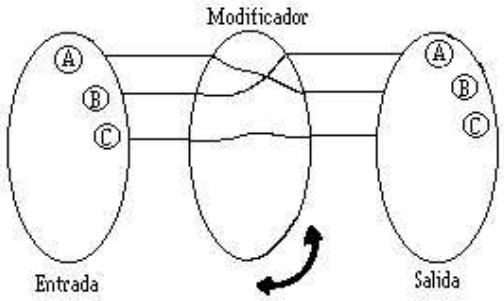


Figura 1.2: Esquema de la Máquina *Enigma*.

Así pues, los alemanes, pese a ser derrotados, manejaban el sistema de codificación más seguro del mundo en aquellos días.

El mencionado grupo de científicos empleaba el que hoy se considera el primer computador —aunque esta información permaneció en secreto hasta mediados de los 70—. Su uso y la llegada del polaco Marian Rejewski tras la invasión de su país natal cambiarían para siempre el curso de la historia. Los polacos se emplearon a fondo en la tarea de romper la codificación de la *Enigma*; con ayuda de la información vendida (planos, claves) a ingleses y franceses por un traidor teutón, Thilo Schmidt, pudieron construir una réplica de la máquina militar (hasta entonces habían trabajado sobre una versión "comercial"), dándose cuenta entonces de una gran Verdad de la Criptografía: que la seguridad depende exclusivamente de la clave, no de conocer el algoritmo.

No obstante, no cejaron en su esfuerzo, y Marian Rejewski llevó a cabo estudios sobre cadenas y periodicidad teniendo en sus manos los libros de claves y conociendo que al principio de cada mensaje los alemanes enviaban la propia clave. Sus trabajos de Criptoanálisis para romper la *Enigma* dieron al fin su fruto: consiguió tabular el espacio de claves hasta reducirlo a la cifra de 105,456. Acto seguido, se desarrolló una máquina que exploraba dicho espacio, cuantitativamente mucho más reducido, por fuerza bruta. Este sistema, que definitivamente rompía con el mito de la *Enigma*, se denominó *Bomba*.

Desde entonces hasta hoy ha habido un crecimiento espectacular de la tecnología criptográfica, si bien la mayor parte de estos avances se mantenían —y se siguen manteniendo, según algunos— en secreto. Financiadas fundamentalmente por la NSA (*National Security Agency*, Agencia Nacional de Seguridad de los EE.UU.), la mayor parte de las investigaciones hasta hace relativamente poco tiempo han sido tratadas como secretos militares. Sin embargo en los últimos años, investigaciones serias llevadas a cabo en universidades de todo el mundo han logrado que la criptografía sea una ciencia al alcance de todos, y

que se convierta en la piedra angular de asuntos tan importantes como el comercio en Internet.

Con el desarrollo de la cultura informática, han surgido nuevas aplicaciones de la Criptología, debido fundamentalmente al manejo de gran cantidad de información. En algunos casos, como en las redes informáticas, dicha información está a disposición de muchos usuarios, lo cual plantea la necesidad de que los datos estén protegidos durante su transmisión y durante su almacenamiento.

Al mismo tiempo, este desarrollo de la informática ha producido un cambio radical en el concepto de seguridad de los sistemas criptográficos, pues aquellos que eran supuestamente seguros frente a procedimientos manuales sucumbieron ante la eficacia de los ordenadores. De esta forma, la supuesta seguridad de los sistemas antiguos o clásicos ha tenido que ser sustituida por una seguridad matemática y computacionalmente demostrable en los sistemas modernos. Además, cuanta mayor protección y seguridad se le quiere dar a la información, mayor es la profesionalización que este campo requiere. También, como se ha podido ver, la evolución se hace más rápida según las necesidades del momento histórico en que nos hallemos. Se incrementan paralelamente los niveles a los que se necesita definir e implementar un mecanismo que garantice dicha seguridad: *físico* (hardware), *lógico* (relacionado con el software) y *administrativo-legal* (leyes penales, leyes administrativas,...).

Muchas son las voces que claman por la disponibilidad pública de la Criptografía. La experiencia ha demostrado que la única manera de tener buenos algoritmos es que éstos sean públicos, para que puedan ser sometidos al escrutinio de toda la comunidad científica. Casos claros de oscurantismo y de sus nefastas consecuencias han sido la caída del algoritmo que emplean los teléfonos GSM en menos de cuarenta y ocho horas desde que su código fue descubierto o los graves problemas de seguridad que presentaba el protocolo de comunicaciones seguras punto a punto que Microsoft incluía en Windows NT. La seguridad no debe basarse en mantener los algoritmos ocultos, puesto que éstos, tarde o temprano, acaban siendo analizados y descritos, sino en su resistencia demostrada tanto teórica como prácticamente, y la única manera de demostrar la fortaleza de un algoritmo es sometiéndolo a todo tipo de ataques. El último capítulo de esta historia ocurrió en el verano de 1999, cuando un programador denunció una supuesta *puerta trasera* en el código criptográfico de todas las versiones de Windows. Como este código permanece en secreto, y se considera delito su análisis —¿qué pensaría usted si se compra un coche y se le prohíbe desarmarlo para ver cómo funciona?—, es desgraciadamente imposible que Microsoft pueda despejar cualquier sombra de duda sobre la seguridad de sus sistemas operativos.

Es imposible desligar la Criptografía moderna de todas las consideraciones políticas, filosóficas y morales que suscita. Recordemos, por ejemplo, que el software criptográfico está sujeto en EE.UU. a las mismas leyes que el armamento nuclear, y que en Europa se pretende elaborar legislaciones parecidas. Algunos gobiernos tienen intención de almacenar todas las claves privadas de sus ciudadanos y considerar ilegales aquellas que no estén registradas. Es como pedirnos a todos que le demos a la policía una copia de las llaves de nuestra casa —con el pretexto, por supuesto, de luchar contra el terrorismo y el narcotráfico—.

Existe un falaz argumento que algunos esgrimen contra el uso privado de la Criptografía, proclamando que ellos nada tienen que ocultar. Estas personas insinúan que

cualquiera que abogue por el uso libre de la Criptografía es poco menos que un delincuente, y que la necesita para encubrir sus crímenes. En ese caso, ¿por qué esas personas que *no tienen nada que ocultar* no envían todas sus cartas en tarjetas postales, para que todos leamos su contenido? o ¿por qué se molestan si alguien escucha sus conversaciones telefónicas? Defender el ámbito de lo privado es un derecho inalienable de la persona, que debe prevalecer sobre la obligación que tienen los estados de perseguir a los delincuentes.

1.2.1. La Criptografía en nuestros días

La Criptografía moderna clasifica sus métodos de la siguiente manera:

$$\begin{cases} \text{Simétricos} & \text{(o de Clave Privada)} \\ \\ \text{Asimétricos} & \text{(o de Clave Pública)} \end{cases}$$

Ejemplos de los primeros (que estudiaremos en capítulos sucesivos) son los algoritmos DES, IDEA, etc. así como los usados en PGP. Entre los segundos, cabe destacar el RSA. La historia del algoritmo **DES**, uno de los más famosos, comienza en 1973 con la organización del primer concurso de Criptografía en EE.UU. Aunque no tiene demasiado éxito, en la convocatoria del año siguiente, en 1974, un empleado de IBM, Feistel, presenta su *método Lucifer*, cuyas ideas se aplican desde entonces a una familia de métodos de cifrado que ha dado en llamarse **métodos de Feistel** (ó basados en).

Poco más tarde, la NSA aprueba el método Lucifer modificado, el origen definitivo del DES, aunque reduce el número de bits de la clave de 128 a 56 (k= $2^{56} = 7'2 \times 10^{16}$ claves posibles). En 1981 adoptará su variante DEA. Otras variantes han sido el *crypt*, utilizado en el cifrado de los passwords en las máquinas Unix, o el *TDES* (por *triple* DES), que simplemente hace 3 "pasadas" del algoritmo, con lo que es más potente y robusto.

El algoritmo **RSA**, basado como todos los métodos asimétricos en el coste computacional de la ruptura del cifrado (tienen un coste aproximado al de una factorización), recibió su nombre de las iniciales de sus creadores: Ronal Rivest, Adi Shamir y Adlemann.

En cuanto a la evolución inmediata de la Criptografía, simplemente indicaremos un nombre que es probable que empiece a oírse cada vez más: el de los *métodos asimétricos*.

1.3. Aspectos Administrativos y Legales

En España, lo referente a la protección y seguridad de la información está regulado por diversas leyes, que mencionamos a continuación:

- Constitución española, el Derecho a la Privacidad.
- Directiva de la Unión Europea 95/46 del 24 de octubre, Protección de Datos de las Personas Físicas, haciendo especial hincapié en el tratamiento de dichos datos.
- Real Decreto 428/1993 del 26 de marzo, aprueba los Estatutos de la Agencia de Protección de Datos (APD).

- ◎ Ley General de Telecomunicaciones 11/998 del 24 de abril, se publica en el BOE del 25 de abril de 1998.
- ◎ Real Decreto 994/1999 del 11 de junio, medidas de seguridad de los ficheros que contengan datos de carácter personal.
- ◎ Real Decreto 14/1999 del 17 de septiembre, sobre la firma electrónica digital (función *md5*) para garantizar la proveniencia de los documentos electrónicos y dar validez legal a los mismos en dicho formato.
- ◎ Ley Orgánica 15/1999 del 13 de diciembre, sobre la Protección de Datos de Carácter Personal (LOPD). Sustituye a la Ley Orgánica 5/1992 del 29 de octubre, conocida como LORTAD. Ambas regulan derechos sobre el honor, la intimidad, cómo garantizar el secreto...
- ◎ Real Decreto 195/2000 del 11 de febrero, tiempo límite para que las empresas en territorio español cumplan lo anterior.
- ◎ Código Penal: se tipifican delitos contra la propiedad intelectual, infidelidad en la custodia de documentos, ataques intencionados a la integridad o confidencialidad, pirateo de software, etc.

Se considera **información de carácter personal**[4] cualquier información concerniente a *personas físicas* identificadas ó identificables mediante el procesado.

Cuando se solicitan datos de carácter personal han de cumplirse una serie de requisitos:

- ⊡ Debe indicarse la existencia de un fichero ó tratamiento, sea físico ó electrónico.
- ⊡ Debe explicarse la finalidad de la recogida de los datos.
- ⊡ Debe informarse del destinatario de la información.
- ⊡ Debe introducirse la identificación y dirección del tratamiento de la información.
- ⊡ Debe especificarse el carácter de los datos (algunos obligatorios, otros no).
- ⊡ Debe notificarse la consecuencia de la negación de la suministración de los datos.
- ⊡ Debe anunciarse la posibilidad de ejercer el derecho a acceder, rectificar y cancelar los datos.

Todo ello es necesario salvo que exista una ley que diga lo contrario, que las fuentes de obtención de los datos sean públicas (guías telefónicas, por ejemplo) o si los fines son considerados históricos, estadísticos o científicos.

También es necesario el consentimiento del afectado si se piensa llevar a cabo un tratamiento de los datos en cuestión, salvo, de nuevo, si una ley lo exceptúa, si la información es recogida por medio de las funciones definidas por las administraciones públicas (siempre que el tratamiento se ciña estrictamente a su ámbito), o bien si se refiere a relaciones laborales o de negocios.

[4]Por tanto, no incluye empresas —salvo que sean sociedades privadas, por ejemplo—.

Se consideran **datos especialmente protegidos** todos los referentes a:

- ▷ Ideología, religión o creencias.
- ▷ Afiliación sindical.
- ▷ Origen racial.
- ▷ Salud.
- ▷ Vida sexual.
- ▷ Infracciones penales y administrativas (multas,...).

Sólo entidades *especializadas* (Iglesia, sindicatos,...), estrictamente relacionadas con estos temas pueden contener información al respecto en sus bases de datos.

Todo profesional informático **debe**, además, **mantener el secreto**, esto es, no difundir a terceros información relativa a los datos que maneja o a los que tiene acceso.

La Agencia de Protección de Datos Española exige, además, que se definan:

Responsable del fichero Es la cabeza visible, el máximo responsable de la seguridad a todos los niveles, fundamentalmente lógico y físico, contra quien penarán las sanciones en pago a cualquier infracción que pueda cometerse. Debe ser una persona estable (no alguien que cambie cada cierto tiempo...), aunque puede asociarse con un cargo. También debe ser una persona de cierto rango.

Responsable de seguridad

Responsable de tratamiento

En general, lo mínimo exigible es:

- ↪ Responsable del fichero.
- ↪ Finalidad de la recogida de los datos.
- ↪ Ubicación de los mismos.
- ↪ Tipo de datos a recoger.
- ↪ Medidas de seguridad adoptadas.

En cuanto a la comunicación de datos entre administraciones públicas, siempre ha de ser notificada, salvo cuando el tratamiento tenga carácter histórico o estadístico o bien cuando el destinatario sea la Agencia Tributaria.

1.3.1. Algunas definiciones

Entre otros, la APD define los siguientes términos:

Seguridad informática Conjunto de recursos de todo tipo (humano, sistemas, procedimientos, normas...) que garantizan la confidencialidad, integridad y disponibilidad de la información.

Incidente de seguridad Hace referencia a cualquier evento que produzca un incidente que afecte a la integridad, disponibilidad o confidencialidad de datos y/o sistemas físicos (ver figura 1.3, página 20). Se habla, respectivamente, de pérdidas lógicas y físicas.

Figura 1.3: Causas de incidentes de seguridad.

Análisis de riesgos Toda organización debe llevar a cabo un análisis de los incidentes que es susceptible de padecer, y definir acciones con vistas a evitarlos (ellos o sus consecuencias) en la medida de lo posible, esto es, de cara a *minimizar* el riesgo de que ocurran.

El análisis de riesgos depende mucho del tipo de organización, y también de otros factores más dispares, como su ubicación geográfica, etc. Asimismo, es importante examinar las causas de los factores de riesgo. De acuerdo con ello, se establecerán normas y procedimientos, o bien se designarán responsables que se hagan cargo de ello, es decir, se acordará una **política de seguridad**.

1.3.2. Legislación vigente

La principal ley a la que debemos prestar atención es al RD 994/1999 del 11 de junio. Este Real Decreto establece medidas organizativas y de índole técnico para garantizar la seguridad de ficheros automatizados, tratando la seguridad en centros, la seguridad de equipos, de sistemas e incluso del software.

Se definen **tres niveles de seguridad** según el tipo de datos que se recojan:

1) Nivel básico.

2) Nivel medio.

3) Nivel alto.

El *nivel básico* hace referencia a los ficheros que contengan datos de carácter personal que no constituyan un perfil de las personas físicas, como por ejemplo:

√ datos de tipo identificativos (nombre, apellidos, DNI...)

√ características personales (altura, color de ojos...)

√ circunstancias sociales (estado civil...)

√ información académica y profesional (expediente, posición...)

√ empleo

El *nivel medio* afecta a información del tipo:

√ servicios financieros

√ solvencia patrimonial y crédito (poder adquisitivo...)

√ infracciones

Por último, el *nivel alto* se refiere a:

√ datos recabados en relación con temas policiales

√ ideología, creencias, religión, origen racial, salud...

En función del nivel de seguridad en que se encuentre un documento, base de datos, etc., deben adoptarse una serie de medidas técnicas concretas. En la práctica, una empresa puede encontrarse con un maremágnum de bases de datos y ficheros en sus sistemas, clasificables en diferentes niveles. En casos así, lo mejor, claro está, es aplicar el nivel más alto a todos ellos.

Medidas de seguridad a establecer

♭ Comprobación de la seguridad en los accesos por red. Se busca que la seguridad del entorno en red garantice la misma seguridad que si todos los accesos se hiciesen de forma local.

♭ Garantizar la seguridad de los datos fuera de su ubicación física.

♭ Garantizar la seguridad de los ficheros temporales y el borrado de los mismos.

♭ Nombrar un responsable (fundamental).

♭ Actuar en el establecimiento de un registro de incidentes o incidencias (que puede a su vez ser registrable en la APD).

Además, según el nivel, deben acatarse una serie de medidas de seguridad concretas:

Nivel básico :

- ⇀ Mantener un listado actualizado de usuarios con acceso a la información.
- ⇀ Establecer mecanismos de identificación y autenticación para el acceso a los datos.
- ⇀ Definir un esquema de usuarios y claves (cifrado), estrategias de renovación periódica (esquemas de caducidad), etc.
- ⇀ Establecer métodos de inventariado, clasificación de soportes de almacenamiento. A este respecto, tener en cuenta la gran problemática de las copias de seguridad (¡muy importante!). Definir un calendario de backups.

Nivel medio :

- ⇀ Definir un calendario de controles periódicos de chequeo del cumplimiento de las normativas de seguridad.
- ⇀ Definir las medidas que se utilizarán a la hora de desechar todo tipo de dispositivos.
- ⇀ Realizar una auditoría de seguridad cada 2 años.
- ⇀ Mecanismos de identificación y autentificación para el acceso a datos restringidos.
- ⇀ Mecanismos para limitar accesos reiterados y no autorizados (evitando así en cierta medida los ataques por fuerza bruta).
- ⇀ Control de acceso físico a los locales.
- ⇀ Procedimientos bien definidos de recuperación de datos, a dos niveles: cómo pueden los usuarios autorizados acceder y recuperar información, y cómo se recupera información ante desastres de todo tipo (terremotos, incendios, errores de dipositivos...).

Nivel alto :

- ⇀ Usar cifrado antes de la distribución de cualquier tipo de soporte que contenga datos de la base en cuestión (es decir, en el envío, ya por red o por mensajería, los soportes han de ir codificados).
- ⇀ Mantener un registro de acceso a la información donde conste al menos el identificador de usuario, fecha, hora, fichero accedido, si el acceso ha sido o no permitido, etc. Conservar un histórico de un período de 2 años.
- ⇀ Realizar un informe de todo el registro de incidencias una vez al mes.
- ⇀ Las copias de seguridad han de residir en lugares diferentes a los del sistema de información.

Cada nivel, por supuesto, incluye todas las medidas del anterior.

Plan de adaptación

El plan de adaptación establecido legalmente consta de 5 fases:

(1) Análisis.

(2) Elaboración.

(3) Desarrollo o Implementación.

(4) Formación.

(5) Auditoría.

que describimos a continuación:

Fase de Análisis :

- ↱ Determinar los ficheros, bases de datos... afectados (que se deben proteger). Las medidas que se redactan a continuación se tomarán *para cada uno de ellos*.
- ↱ Comprobar en la APD que dichos archivos no figuran y qué nivel de seguridad se tiene para cada uno.
- ↱ Comprobar la seguridad de los ficheros a través de la red.
- ↱ Comprobar la seguridad de los ficheros en su lugar de ubicación física.
- ↱ Comprobar la existencia (y en su defecto definirlo) de un responsable por fichero (y comunicarlo pertinentemente).
- ↱ Comprobar la existencia de un registro de incidencias.
- ↱ Comprobar la existencia de un registro de usuarios, de mecanismos de identificación, de una política de renovación de claves...
- ↱ Comprobar el inventario y clasificación de soportes.
- ↱ Documentar el inventario y comprobar y documentar la existencia de copias de seguridad (¿política de los backups: incremental...?).
- ↱ Definir un calendario de controles periódicos encaminados a comprobar y verificar que se cumple la normativa.
- ↱ Definir las medidas a tomar al desechar o reutilizar un dispositivo.
- ↱ Comprobar la realización de auditorías cada dos años.
- ↱ Comprobar los mecanismos de identificación de usuarios.
- ↱ Comprobar que existen mecanismos de control de intentos de acceso infructuosos reiterados.
- ↱ Comprobar las medidas de seguridad físicas (control de acceso a los edificios).
- ↱ Comprobar la E/S de soportes informáticos.
- ↱ Comprobar la existencia de procedimientos de recuperación de datos (por ejemplo ante desastres, ataques...)
- ↱ Comprobar los procedimientos de cifrado de los datos.
- ↱ Comprobar la existencia de registros de acceso.

- Comprobar la realización de informes mensuales.
- Comprobar la realización de copias de seguridad y su ubicación en diferentes lugares físicos.

Fase de Elaboración Debe recogerse en un *documento de seguridad*:

- Ámbito de aplicación y especificación del mismo.
- Recursos protegidos.
- Medidas, normas, procedimientos y reglas adoptadas para garantizar la seguridad.
- Funciones y obligaciones del personal.
- Definición de la estructura de los ficheros.
- Descripción de los sistemas de información donde residen los datos.
- Procedimiento de gestión de incidentes (de la forma: suceso ⇒ acciones...).
- Copias de seguridad.
- Información sobre responsables (identificación...).

Fase de Implementación Han de llevarse a cabo las normas de seguridad documentadas en la fase anterior. Se incide en la configuración de aplicaciones, sistemas operativos, firewalls, arquitecturas de red... Ello puede significar la modificación de las técnicas que se venían usando habitualmente, incluida la propia estructura de la organización, en el peor de los casos.

Fase de Formación Se recomienda dar al personal informático formación orientada a:

- conocimiento de mecanismos orientados al control de acceso a dispositivos, a la información...
- gestión de soportes
- registro de incidencias (cómo se notifican, cómo se tratan...)
- identificación y autenticación
- copias de seguridad, respaldo...

Fase de Auditoría Su realización es necesaria en los niveles medio y alto. Consiste en llevar a cabo un análisis de:

- la red de comunicaciones.
- los sistemas operativos (mejor homogeneizar).
- los ficheros automatizados, bases de datos...
- los mecanismos de acceso remoto
- a nivel físico (UPS, alimentación, sistemas contraincendios, sistemas de acceso a edificios...)

Tras ello se realiza una identificación de puntos débiles y en función de éstos se dan recomendaciones para asegurarlos. Por último, puede crearse un manual para navegantes[5].

[5]Hay herramientas de metodología para todo esto.

1.4. Vulnerabilidades de los Sistemas Informáticos

Los puntos débiles de los sistemas informáticos se clasifican en tres áreas fundamentales:

- **Nivel hardware**.
- **Nivel software**.
- **Nivel de datos**.

En cada una de ellas se recomiendan una serie de medidas de seguridad particulares y adecuadas, que vemos a continuación.

1.4.1. Medidas generales a nivel hardware

A fin de mantener la seguridad del hardware, es fundamental actuar en las siguientes líneas:

√ Aislamiento de CPD's.

√ Sistemas contra incendios.

√ Sistemas de alimentación de corriente ininterrumpida (UPS, generadores,...).

√ Hardware redundante (RAID —por hardware, por software—, mirror de datos, dobles controladoras,...).

√ Mantenimiento.

1.4.2. Medidas generales a nivel software

Con respecto a la seguridad relacionada con el software, es clave:

√ La selección del sistema operativo (se recomienda la mayor homogeneización de sistemas posible).

√ Parcheado de sistemas operativos.

√ Parcheado de programas.

1.4.3. Medidas generales a nivel de datos

En general, toda la presente asignatura se ocupa mayoritariamente de este punto, pues este aspecto gira fundamentalmente en torno a la **criptografía**.

El conjunto de medidas adoptado por una organización en cada uno de los mencionados niveles define su **política de seguridad**.

Pese a todo, debemos tener muy claro que el **sistema seguro no existe**, no existe un sistema de información sobre el que no se conozcan tipos de ataques que puedan vulnerarlo, ya sea mediante el tipo de ataques que hemos visto, o mediante otro tipo de

agresiones que comentaremos también en su momento, como *bombas lógicas*, *troyanos*, *gusanos*, *virus*,...

$$S.I. = f(\$, \text{eficiencia})$$

De todos modos, normalmente el incremento de las medidas de seguridad va ligado con una disminución de la flexibilidad de movimientos del usuario, algo que también debemos tener en cuenta.

1.5. Criterios de Evaluación de la Seguridad

Existen una serie de criterios para evaluar la seguridad de un sistema o conjunto de sistemas:

— TCSEC (*Trusted Computer System Evaluation Criteria*), más conocido como "el libro naranja" de los EE.UU.

— ITSEC (*Information Tecnology Security Evaluation Criteria*).

— Otros criterios europeos.

Todos ellos definen la necesidad de garantizar la *confidencialidad* y el *secreto* (la información debe estar disponible sólo para los usuarios autorizados[6] a manejarla), la *integridad* (asegurar que la información no se ha falseado; se usan herramientas como la firma digital) y la *accesibilidad* (garantía de quién, cómo y cuándo accede a la información). También se añaden la *autenticidad* (verificación del origen y el destino de la información) y la *imposibilidad de rechazo* o *no repudio* (que cualquier entidad que envíe o reciba información no pueda negar el hecho de haberla enviado o recibido).

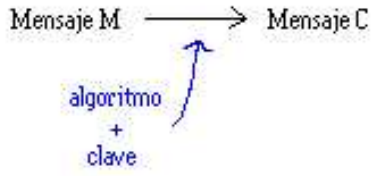
1.6. Medidas de Seguridad. Niveles

Existen una serie de niveles sobre los que son susceptibles de aplicarse medidas de seguridad:

No existe un mecanismo que atienda a todos ellos, pero sí un *conjunto de mecanismos*. Casi todos ellos hacen uso de la criptografía. Algunos son:

[6] Un *usuario* no es sólo una persona, puede ser un programa —threads—, un entorno,...

- Intercambio de autenticación (garantiza que las entidades origen y destino son las que deben ser).
- Cifrado. Dos mundos: simétrico (cifradores de bloques, cifradores de flujo) y asimétrico.



- Integridad de datos, ICV (*integrity check value*): se incluye un valor de comprobación de identidad (el caso más simple son los *checksums*).
- Firma digital.
- Control de acceso (sistemas de contraseñas, listas de acceso o ACL's, firewalls en hosts y routers...).
- Inclusión de tráfico de relleno.
- Control de encaminamiento, empleo de rutas redundantes.
- Unicidad: inclusión de número de secuencia, fecha, hora, etc. en los paquetes, para evitar reactuación y resecuenciación.
- Gestión de red.
- Gestión de seguridad, tanto su generación, localización, distribución y control de acceso a la información secreta, como la política de servicios, mecanismos de seguridad, detección de infracciones y acciones a realizar en caso de detección de anomalías.
- Política de gestión de claves: generación (no débiles, largas, aleatorias), distribución, almacenamiento, tiempo de vida, destrucción, aplicación. Hay que tener en cuenta que:
  - La elección de un espacio de claves reducido facilitará la penetración del sistema por fuerza bruta; cualquier elección pobre lo hará vulnerable a herramientas de chequeo de claves, pruebas con diccionarios[7], etc. Lo más seguro son siempre claves aleatorias, aunque el problema sigue existiendo: ¿cómo memorizarlas? ¿almacenarlas? Para ello hay soluciones que utilizan una *frase de paso* que genera la clave del usuario (estrategia que usa PGP).
  - En la distribución de claves puede optarse por un intercambio directo o por utilizar un *Centro de Distribución de Claves*, que en ocasiones incluso las asignan[8]

[7] Del estilo de *John the ripper*, *Satan*,...

[8] Debido a la necesidad del intercambio de claves para comunicaciones cifradas o intercambios de documentos encriptados, cuando hay redes involucradas siempre es mejor utilizar métodos asimétricos, métodos de clave pública, de los que ya hablaremos.

- Alternativas muy seguras para el almacenamiento de claves son las tarjetas de banda magnética o tarjetas inteligentes.
- Otro factor muy importante en la seguridad es el tiempo de vida: cuanto mayor sea, más inseguro es el esquema. Algunos protocolos orientados a conexión utilizan una clave distinta en cada conexión; los no orientados a conexión suelen hacer uso de claves temporales, válidas por un período limitado de tiempo.

1.7. Fundamentos de Criptología

Como ya hemos mencionado alguna vez, la **Criptología** es la ciencia que se encarga de todo la relacionado con la *Criptografía* y la **Criptografía** es la técnica o ciencia que permite proteger la información por medio del uso del *cifrado* de forma que solo pueda ser descifrado por el remitente. El **cifrado** es la técnica que permite, a partir de un mensaje en claro aplicando un algoritmo y una clave que tenga inversa, encriptarlo de forma que sea muy difícil su descifrado careciendo del conocimiento tanto la clave como del algoritmo empleados.

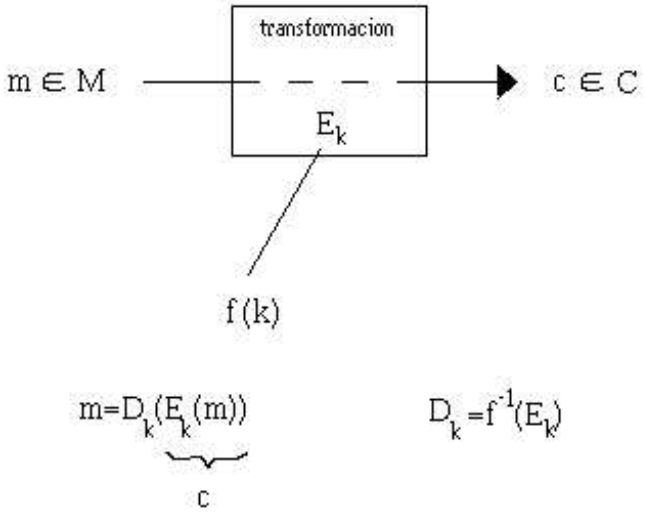


Figura 1.4: Componentes de un Criptosistema.

Los *Criptosistemas* pueden implementarse tanto en hardware como en software. La diferencia es que los primeros son, evidentemente, más rápidos.

Los ataques, como ya hemos visto, pueden producirse:

1. A partir de texto cifrado (obteniendo la clave; cuanto más texto se tenga, más fácil es).
2. A partir de algún mensaje conocido.
3. Elección de mensajes (ver figura 1.5).

Figura 1.5: Criptoanálisis.

Todo **sistema criptográfico** o **criptosistema** consta de:

√ Espacio de mensajes $M = \{m_1, m_2, \ldots\}$.

√ Espacio de textos cifrados $C = \{c_1, c_2, \ldots\}$.

√ Espacio de claves $K = \{k_1, k_2, \ldots\}$.

√ Una familia de transformaciones de cifrado $E_k : M \to C \quad k \in K$.

√ Una familia de transformaciones de descifrado $D_k : C \to M \quad k \in K$.

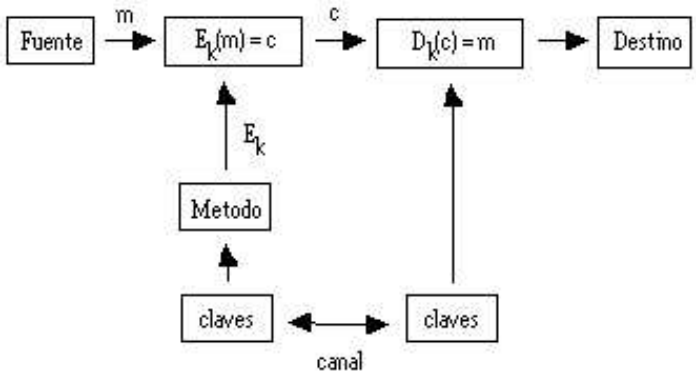


Figura 1.6: Sistema criptográfico.

En la actualidad se distinguen dos tipos de criptosistemas: **simétricos** o de clave privada (DES, IDEA, RC2, RC4, etc) y **asimétricos** o de clave pública (RSA, DSS, Diffie & Hellman,...).

Shanon midió el *secreto* de un cifrador en función de la *incertidumbre* sobre el texto en claro. Estudió a fondo los sistemas simétricos, y formuló la fundamentación de la Teoría

de la Comunicación sobre la Teoría de la Información, la Teoría de Números y la Teoría de la Complejidad Algorítmica.

Utilizó la noción de **entropía** (grado de inestabilidad), la de **redundancia** de un lenguaje (cantidad de información, de cosas sin sentido que se pueden escribir en él), algo que emplearía después Hellman para esbozar su método (el *método de Hellman*: función(redundancia)=lenguaje; se han propuesto **técnicas de difusión y confusión** que intentan romper la frecuencia de los criptogramas).

1.8. Métodos Criptográficos

Esta sección expone a grandes rasgos los más importantes de los métodos criptográficos conocidos.

1.8.1. Clásicos

Estudiaremos en primer lugar los métodos que dieron lugar al origen de la Criptografía, los llamados **métodos clásicos**, que se basan principalmente en la sustitución y transposición de caracteres.

Sustitución Simple Monoalfabeto

Este método de cifrado consiste en sustituir cada letra o carácter del mensaje (texto en claro) por otra letra que forma parte del cifrado (texto cifrado).

Para efectuar esta sustitución existen varios métodos, los cuales en general se pueden expresar mediante la transformación:

$$E(m) = (a \cdot m \pm b) \mod n$$

donde

m es el valor numérico asociado a cada letra del mensaje

E es la función de cifrado

E(m) es el valor numérico asociado a la letra correspondiente del mensaje cifrado

a es una constante que determina el intervalo de separación entre dos letras del cifrado cuando éstas son consecutivas en el alfabeto del mensaje original. Para que el alfabeto de letras equivalentes sea completo, es decir, que no se repita ninguna letra (que no ocurra que dos letras sin codificar den lugar a la misma al codificarlas), es preciso que sea primo con n

b es una constante que determina el desplazamiento entre la correspondencia de las letras del cifrado y del mensaje

n es el número de letras del alfabeto

La transformación mencionada es conocida como método de sustitución simple con el nombre de **método César**.

Determinada la equivalencia entre una letra del mensaje y del cifrado, queda determinada la clave, por lo que el sistema no ofrece una gran seguridad.

En el criptoanálisis de la **sustitución monoalfabeto tipo Cesar** se ha de tener en cuenta que el alfabeto equivalente empleado es uno desplazado con respecto al original un número entero b, algo que, como decimos, lo hace fácilmente vulnerable. Se descifra utilizando, por ejemplo, un análisis de frecuencias de las letras en el texto cifrado, comparándolas con las frecuencias características en el alfabeto original, hasta encontrar el acoplamiento o correspondencia adecuada.

Sustitución Homofónica

Esta sustitución establece una aplicación entre el conjunto de caracteres del alfabeto de mensajes $\mathrm{A_M}$ y el conjunto potencia (conjunto de los posibles subconjuntos) de los elementos del alfabeto cifrado $\mathrm{A_C}$:

$$\mathrm{E} : \mathrm{A_M} \rightarrow 2^{\mathrm{A_C}}$$

Dado que el conjunto potencia $2^{\mathrm{A_C}}$ comprende los posibles subconjuntos que se pueden formar con los elementos de $\mathrm{A_C}$, a los elementos de cada subconjunto en que se puede cifrar un solo caracter se les denomina **homófonos**. El método pretende destruir la frecuencia de los caracteres del mensaje.

El mensaje $\mathrm{M}=\mathrm{m_1 m_2 \ldots m_n}$ se cifra como $\mathrm{C}=\mathrm{E(m_1)E(m_2)\ldots E(m_n)}$ donde $\mathrm{E(m_i)}$ se toma aleatoriamente del subconjunto de homófonos de $\mathrm{m_i}$, siendo necesario no repetir la sustitución mientras no se hallan utilizado todos los homófonos.

Sustitución Polialfabeto

Este método de sustitución consiste en el uso de varias sustituciones simples en el cifrado de un mensaje. Para ello se utiliza una palabra clave cuyas letras definen los desplazamientos de los diferentes alfabetos equivalentes en sustituciones del tipo César, que se aplican a las letras del mensaje original.

Esta transformación (cifrado de Vigenére) se puede representar mediante:

$$E = E(m_j) = (m_j + k_i) \mod n$$

donde k_i es el desplazamiento de cada letra de la palabra clave respecto al alfabeto base. El valor máximo del subíncide i es la longitud de la palabra clave, que constituye el período de la sustitución polialfabeto denotado por la letra d.

Puesto que la sustitución simple utiliza una aplicación biyectiva entre las letras del cifrado y del mensaje, la frecuencia de las letras permanece en el texto cifrado. En la sustitución polialfabética al sustituirse una letra del texto en claro por diferentes letras en el texto cifrado se pretende distruir las frecuencias características de las letras del texto original.

Cifrado de Beaufort.- El cifrado de Beaufort utiliza la función de cifrado

$$E = E(m_j) = (k_i - m_j) \mod n$$

siendo la función del descifrado

$$D = D(c_j) = (k_i - c_j) \mod n$$

Con ello, este cifrado invierte las letras del alfabeto equivalente y las desplaza (k_i+1) posiciones. Una variante del cifrado de Beaufort es la que cifra según la transformación

$$E(m_j) = (m_j - k_i) \mod n$$

que es equivalente al cifrado de Vigenére con clave (n-k_i).

Autoclave.- Una variante del de Vigenére es el cifrado de **autoclave** donde el texto en claro se utiliza como clave, empleándose igualmente una clave primaria. En el descifrado, se toma como clave la primaria seguida del texto en claro que se va descifrando con ella y concatenando el resto del texto en claro que se va obteniendo:

```
ENVIASUMINISTROS   (texto en claro)
JUEVESENVIASUMIN   (clave=JUEVES)
NHZDEKYZDVALNDWF   (texco cifrado)
JUEVESENVIASUMIN   (clave)
ENVIASUMINISTROS   (texto en claro)
```

Cifrados con clave continua.- Son cifrados en los que la longitud de la clave es tan grande como el mensaje, al objeto de evitar que la clave se repita. Para ello, se puede utilizar como clave una secuencia aleatoria no repetitiva. Entre otros, se puede pensar en usar como clave el texto de un libro, y cifrar mediante sustitución polialfabeto de acuerdo con los desplazamientos de cada palabra del texto.

No obstante, pese a ser la sustitución polialfabeto más segura que la monoalfabeto, no es inmune al criptoanálisis. El procedimiento consiste en determinar el número de alfabetos empleado (o periodo, longitud de la clave), separar el texto en partes que fueron cifradas con el mismo alfabeto equivalente y resolver cada parte por separado como si hubiera sido cifrada con un solo alfabeto.

Existen dos maneras de descubrir el periodo, una es el **método de Kasiski** y otra el estudio del **Índice de Coinicidencia**:

Método de Kasiski

Este método se basa en la repetición de grupos de letras y palabras en el lenguaje natural (por ejemplo, en español, -*as*, -*os*, -*es*, -*ción*, *en*-, *co*-, *in*-, *con*, *de*, *y*, *a*, etc.).

Si un texto se cifra con n alfabetos de forma cíclica ($|\text{clave}| = n$), y si una palabra o grupo de letras aparece k veces en el texto en claro, será cifrado aproximadamente $frackn$ veces con el mismo alfabeto. Así, por ejemplo, si se utiliza una palabra clave de 3 letras se podrá disponer en 3 formas diferentes sobre el texto en claro, con lo que, si una palabra o grupo de letras se repite más de 6 veces debe ser cifrada al menos dos veces con

la misma posición de la clave, y esas ocurrencias serán grupos de letras cifradas de la misma forma.

Veamos un ejemplo:

```
DESCONFIANZA CON LOS PRISIONEROS      k=GIN (3 alfabetos)
GINGINGINGIN GIN GIN GINGINGINGI
JMFIWZLPNSHN IWZ QWF VZUYPBSMEAU
   ---       ---
```

Se observa que `IWZ` se repite a partir de las posiciones 4 y 13. Si se calcula la diferencia, se determinan como posibles periodos los factores de ésta: $13 - 4 = 9 \rightarrow 1, 3$ y 9.

Índice de Coincidencia

El índice de coincidencia es una herramienta criptoanalítica que permite evaluar la similitud entre la distribución de la frecuencia de las letras de un texto cifrado y la de la frecuencia característica de las letras de un lenguaje natural. Así, éste mide la varianza de la frecuencia de las letras.

La medida de la dispersión o varianza es:

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

donde p_i es la probabilidad de ocurrencia de *cada letra* en el texto original y n el número de letras del alfabeto.

Desarrollando, se llega a que MD puede escribirse:

$$MD = \sum_{i=1}^{n} {p_i}^2 - \frac{1}{27} = \sum_{i=1}^{n} {p_i}^2 - 0'037 = 0'075 - 0'037 = 0'038$$

(para $n = 27$, castellano; $0 \leqslant MD \leqslant 0'038$). El término $\sum {p_i}^2$ se define como **índice de coincidencia** y puede expresarse como la probabilidad de que dos caracteres en el texto cifrado sean iguales. La siguiente expresión permite medir este índice en base a las frecuencias f_i observadas de cada letra, en un texto de n letras, simplemente estableciendo la relación entre los pares de letras iguales encontrados y los pares de letras posibles en el texto de n letras:

$$IC = \sum_{i=1}^{n} {p_i}^2 = \frac{\sum_{i=1}^{n} f_i(f_i - 1)}{N(N-1)}$$

El índice de coincidencia para cifrados polialfabeto de período p y texto de n letras es:

$$IC = \frac{f_i(f_i - 1)/2}{n(n-1)/2} = \frac{f_i(f_i - 1)}{n(n-1)}$$

De donde

$$MD = IC - 0'037$$

Según esta fórmula, para un texto de $n = 1000$ letras es castellano se obtiene la tabla:

P	**IC**
1	0′075
2	0′055
3	0′049
.	.
10	0′040
grande	0′037

Cuadro 1.3: Periodo e Índice de Coincidencia en Sustituciones Polialfabeto.

Otros tipos de cifrado

Cifrados tipo Vernam.- En la idea de encontrar cifrados perfectos mediante el uso de claves no repetitivas de tamaño igual al del mensaje (*cifrados de clave continua*), Gilbert Vernam de AT&T basó el diseño de un método inmune a la mayoría de los ataques criptoanalíticos. El procedimiento consiste en combinar una larga secuencia aleatoria de números no repetitiva con el mensaje en claro. Si la clave no se repite ni reutiliza, el texto cifrado no muestra la estructura de la clave.

Cuando la información que se cifra es binaria, la combinación de la clave y texto en claro se hace mediante la operación **or exclusivo**, donde la clave consiste en una ristra no repetitiva de dígitos binarios. El **or exclusivo** sirve tanto para el cifrado como para el descifrado, ya que:

$$E(m_i) = m_i \oplus k_i$$

$$E(m_i) \oplus k_i = m_i \oplus k_i \oplus k_i$$

Para la operación de descifrado en el extremo receptor se precisa, pues, disponer de la misma secuencia de dígitos binarios de la clave.

Sustitución Poligráfica.- En este método de sustitución, en lugar de sustituir una letra por otra, permite sustituir digramas, trigramas, etc., de letras, al objeto de destruir las frecuencias de los monogramas cifrando un n-grama de una vez, con lo que se consigue una mayor seguridad. Son ejemplos:

Cifrado Playfair.- Es un cifrado de sustitución digrámica en el que la clave viene dada por una matriz 5×5 de caracteres en las que se distribuyen 25 letras del alfabeto. Se recurre al uso de una palabra clave sin letras repetidas, que se sitúa en las primeras filas del cuadro. A continuación de ella se sitúan el resto de letras del alfabeto que no están en la palabra clave.

Las reglas para cifrar 2 caracteres m_1m_2 del texto en claro son:

1. Si m_1 y m_2 están en la misma fila, c_1 y c_2 son los situados a su derecha.
2. Si m_1 y m_2 están en la misma columna, c_1 y c_2 son los situados a continuación en la columna (de arriba a abajo, circularmente).
3. Si m_1 y m_2 están en distinta fila y columna, se toman los de la diagonal opuesta:
$$\begin{matrix} m_1 & c_1 \\ m_2 & c_2 \end{matrix}$$
4. Si m_1=m_2 se inserta un carácter sin significado entre ambos.
5. Si hay un número impar de letras, se añade al final del texto en claro otra sin significado.

Cifrado Hill.- Realiza el cifrado de k caracteres del mensaje en k caracteres del texto cifrado, mediante el uso de una transformación lineal.
En el caso de sustitución de digramas, la transformación que se realiza es la siguiente:

$$c_1 = (a_{11}m_1 + a_{12}m_2) \mod n$$

$$c_2 = (a_{21}m_2 + a_{22}m_2) \mod n$$

donde

$$K(a) = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

ha de tener inversa ($A \cdot A^{-1} \mod n = I$) para deshacer la codificación $C = (A \cdot M) \mod n$.
Se puede vulnerar este método si se conocen dos digramas cifrados y sus correspondientes textos de mensajes en claro, ya que:

$$M_1 = (m_1, m_2) \qquad M_2 = (m_3, m_4)$$

$$C_1 = (c_1, c_2) \qquad C_2 = (c_3, c_4)$$

$$M = \begin{pmatrix} m_1 & m_3 \\ m_2 & m_4 \end{pmatrix} \quad C = \begin{pmatrix} c_1 & c_3 \\ c_2 & c_4 \end{pmatrix}$$

$$C = A \cdot M \mod n \Rightarrow A = C \cdot M^{-1} \mod n$$

Transposición.- En este método de cifrado no se efectúa ninguna sustitución de letras en el mensaje (texto en claro), sino que se cambia su posición dentro del mensaje. Estos métodos no evitan la aparición de las letras con su frecuencia característica en el texto cifrado y son fácilmente destruibles mediante anagramación o distribución en bloques del texto cifrado. Existen diversos procedimientos para reordenar las letras del mensaje y obtener el mensaje cifrado: *permutaciones*, *posicionamiento en zig-zag*, *distribución en figuras geométricas*,...

Métodos aritméticos.- Existen diferentes tipos:

Adición y sustracción.- Puesto que las operaciones de suma y resta son transformaciones que poseen una inversa, éstas pueden emplearse como métodos criptográficos:

$$E_k(m) = m \pm k \Rightarrow M = E_k(m) \pm k$$

Multiplicación y división.- De la misma forma que con las operaciones de suma y resta, se pueden cifrar mensajes utilizando transformaciones de multiplicación y división (si bien éstas expanden y reducen la longitud del mensaje). En general, puede usarse cualquier función aritmética que tenga inversa.

Conversión de la base del sistema de numeración.- Se puede cifrar la representación numérica entera de un mensaje, ya que la conversión o reconversión son transformaciones reversibles. Con este método, igual que el anterior, la longitud del mensaje puede reducirse o expandirse en virtud de la base del nuevo sistema de numeración.

Estos métodos son sencillos de implementar con computadoras, presentando una gran seguridad, ya que destruyen los parámetros estadísticos del lenguaje, pues a partir del mensaje cifrado o resultado de la operación es difícil determinar los operandos involucrados, es decir, clave y mensaje original. Por otra parte, cualquier error en las operaciones puede ser un error irrecuperable.

Transformaciones lógicas booleanas.- Es posible utilizar operaciones lógicas booleanas como transformaciones criptográficas. No obstante, de todas las posibles operaciones lógicas del álgebra de Boole, sólo la *negación*, la *equivalencia* y el *or exclusivo* poseen operaciones inversas. Todas ellas son además sus propias inversas. El cifrado tipo Vernam es un caso de uso de la transformación booleana *or exclusivo*.

Transformaciones matriciales.- El uso de operaciones matriciales proporciona tranformaciones criptográficas, aunque laboriosas, bastante seguras, pues normalmente destruyen los parámetros estadísticos del lenguaje.

En este método, el mensaje original se transforma mediante un código binario en una sucesión de ceros y unos. A continuación, se disponen esos bits en una matriz de r filas y s columnas, que se suma o multiplica por otra matriz clave, obteniendo la matriz cifrada C:

$$(C) = (M) + (K) \quad \text{ó} \quad (C) = (M) \times (K)$$

La operación de descifrado es posible debido a que la suma y multiplicación de matrices poseen operaciones inversas bajo ciertas condiciones (la suma precisa equidimensionalidad y el producto, que (K) posea inversa única, esto es, que sea cuadrada y no singular). Para reducir el cálculo es conveniente que el número de columnas s sea inferior al de filas r en la matriz (M), ya que al multiplicar las dos matrices el número de multiplicaciones es rs^2 y el de sumas $rs(s-1)$.

En el caso más elemental, la matriz (K) es una matriz ortogonal y su inversa es su traspuesta. Las *matrices de Hadamard* son útiles en este sentido, pues al ser sus elementos de valor ± 1 tienen una inversa que es la traspuesta dividida por el orden de la matriz.

Clasificación de los Métodos Clásicos

A continuación se muestra una figura que resume en un esquema la clasificación de los métodos criptográficos más importantes de los citados anteriormente. Aunque no se han

descrito, se citan los de tipo mixto. Se entiende por sustitución monoalfabética o polialfabética mixta cuando la aplicación entre las letras del alfabeto original y el equivalente se establece de una forma aleatoria.

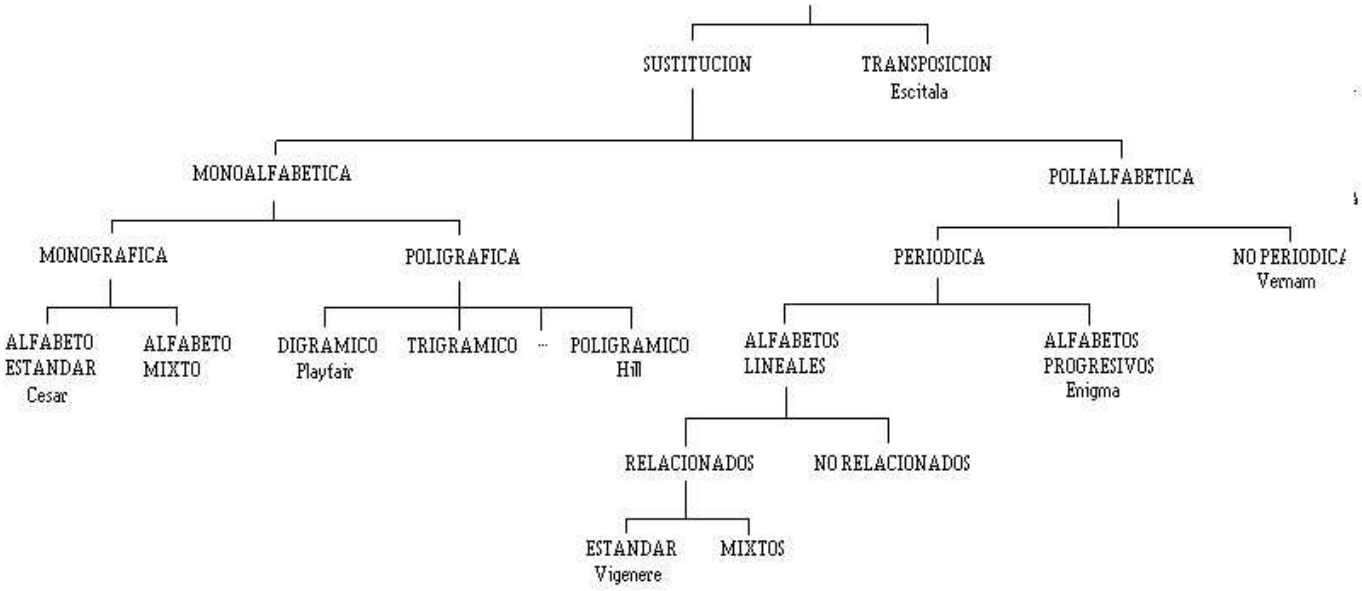


Figura 1.7: Clasificación de Métodos Clásicos

1.8.2. Modernos

Los **Métodos Criptográficos Modernos** están orientados al cifrado de bases de datos, al soporte de cifrado de binarios, etc. Son métodos más seguros y fiables gracias a su filosofía de publicación.

Se clasifican en dos grandes grupos:

- **Criptosistemas de Clave Privada** o **Simétricos**
- **Criptosistemas de Clave Pública** o **Asimétricos**

Simétricos

Los *criptosistemas de clave privada* se caracterizan por utilizar para el cifrado y descifrado la misma clave, que se mantiene secreta. Estos sistemas de clave única son denominados de "clave privada" en contraposición a los de "clave pública" en los que parte de la clave se da a conocer. Por otra parte, por usar la misma clave secreta se denominan también "simétricos", en contraposición a los de clave pública denominados "simétricos". La autenticidad se consigue al permanecer secreta la clave, y por consiguiente, sólo el emisor legítimo puede producir un cierto mensaje cifrado, que puede a su vez ser descifrado por el receptor haciendo uso de la clave compartida.

Estos sistemas presentan, no obstante, una serie de **inconvenientes**:

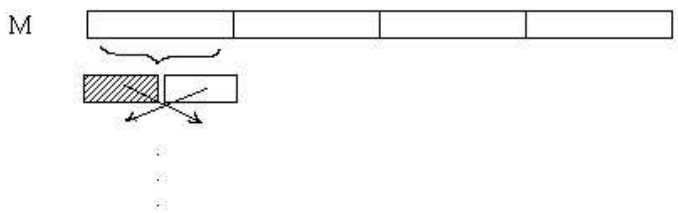
- En cuanto a la *distribución de las claves*, ya que éstas deben intercambiarse con seguridad absoluta.
- En cuanto a la *gestión del número de claves*, ya que aumenta con el cuadrado del número de usuarios que intercambian información secreta en el sistema (si tenemos n usuarios y n intercambios, se trata de $\frac{n(n-1)}{2}$ claves), lo que obliga a limitar el número de participantes en el mismo[9].
- En cuanto a las *firmas digitales*, puesto que no hay forma de incluirlas en un sistema simétrico.

Por todo ello, suelen utilizarse en aplicaciones sencillas y de ámbito local, como pueden ser el `login` y el `password`, que se ejecutan en la propia máquina. Los problemas que se encuentran en los criptosistemas de clave privada son elegantemente solventados por los sistemas de clave pública.

Los sistemas de cifrado simétricos se clasifican a su vez en dos subgrupos, que ha hemos mencionado alguna vez: *cifradores de flujo* y *cifradores de bloques*. Entre los primeros encontramos algunos de los algoritmos de cifrado de clave privada más conocidos, como el DEA (*Data Encryption Algorithm*), su estándar el DES (*Data Encryption Standard*), el IDEA (*International Data Encryption Algorithm*),...

Los sistemas simétricos suelen ser cifradores por bloques, donde M se trocea y se trabaja independientemente con cada uno de los bloques que se obtienen. Suelen aplicar técnicas de sustitución, transposición, suma, resta, multiplicación, división, cambios de base, or exclusivo,... mezcladas y aplicadas en varios ciclos, varias veces. Por ello son, obviamente, algorítmicamente más complejos comparados con los métodos clásicos.

Esta idea fue desarrollada por **Feistel**, e implementada por primera vez en el dispositivo **Lucifer** (1977). Después ha seguido utilizándose, de suerte que los algoritmos de cifrado que se basan en ella se denominan *basados en Feistel*. El procedimiento es sencillo: el bloque de datos a cifrar se divide a su entrada en trozos. Cada uno de ellos, se subdivide a su vez en dos, de los cuales uno se somete a una transformación y el otro no. Posteriormente se intercambian, procediendo así con todos los bloques generados un determinado número de veces.



Un ejemplo:

```
M1 = STAR TREK LA MISION CONTINUA     (sust. con despl. 1c.)
                                      (bloques 4 c.)
```

[9]En un sistema asimétrico, sin embargo, en la misma situación se tendrían solamente $2n$ claves.

```
S1 = TUBS TREK MBNJ SION DPOU NUA
P1 = BUST TREK JNBM SION UOPD NUA
M2 = TREK BUST SION JNBM NUA UOPD
```

Puede incorporarse también una clave en la función de transformación, y según cómo se aplique se tienen:

↪ *Cifradores en cascada*, que usan una clave que en realidad son varias más pequeñas, una para cada iteración.

↪ *Cifradores producto*, donde se tiene una clave a partir de la cual el propio algoritmo genera varias para aplicar en los diversos ciclos[10].

Entre las características de este método están que letras iguales no entregan criptogramas iguales (en el ejemplo anterior, la `A` puede dar `E`, `I` y `D`), y que en M y C no aparecen los mismos caracteres.

Cuadro 1.4: Comparativa entre diferentes sistemas de cifrado simétricos.

Método	**Tamaño bloque**	**Tamaño clave**	**Vueltas**	**Tipo Feistel**
LUCIFER	128 bits	128 bits	16	Sí
DES	64 bits	64 bits[a]	16	Sí
LOKI	64 bits	64 bits	16	Sí
IDEA	64 bits	128 bits	8	Sí/No[b]
SCRIPTJACK	64 bits	80 bits	32	No se sabe

[a] 56 bits de clave + 8 bits de paridad.
[b] Puede serlo o no.

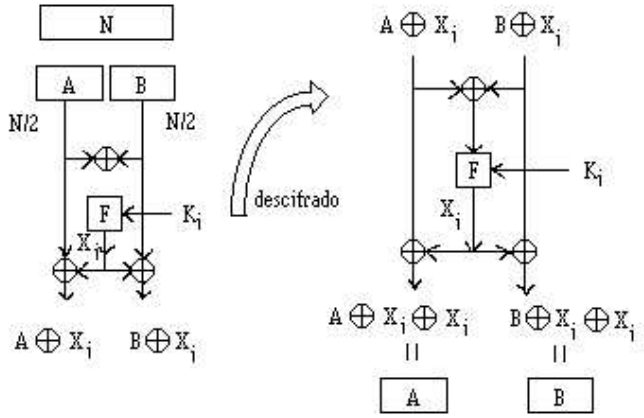


Figura 1.8: Funcionamiento de un cifrado tipo Feistel.

[10] Un comportamiento de este estilo exhibe el sistema de cifrado DES.

El algoritmo **DES** (*Data Encryption Standard*) fue desarrollado en IBM como continuación del dispositivo Lucifer (1974), y ha sido tomado como algoritmo estándar de cifrado por la NSA (*National Security Agency*) americana (1977)[11]. En 1981 el ANSI (*American National Standard Institute*) adopta DES como DEA, evitando hacer uso de la palabra "standard". En la actualidad, ha sido también considerado como posible norma por ISO bajo esta misma denominación.

Este algoritmo fue implementado en hardware y se le dio la categoría de armamento estratégico, prohibiéndose su comercialización en Europa. En síntesis, el sistema cifra un bloque de 64 bits de texto en claro en un bloque de 64 bits de texto cifrado. Para ello usa una clave externa de 64 bits (8 bytes, por lo cual cada clave puede asociarse a 8 caracteres ASCII) en los que los bits de las posiciones octavas de cada byte (8, 16, 24, 32, 40, 48, 56, 64) son bits de paridad impar.

El algoritmo consta de 16 pasos o iteraciones. En cada una de ellas se usan operaciones de or exclusivo, permutaciones y sustituciones. Las permutaciones son de tres tipos: simples, expandidas (en las que se duplican bits) y restringidas (en las que se eliminan ciertos bits). Las sustituciones son no lineales. A grandes rasgos lo que se hace es:

1. Se aplica una **permutación** conocida como **P1** a todo el bloque de 64 bits, simplemente para "descolocar" un tanto el mensaje en claro. Dado que tampoco aporta demasiado, algunas versiones que han dado lugar a otros algoritmos suprimen este paso.

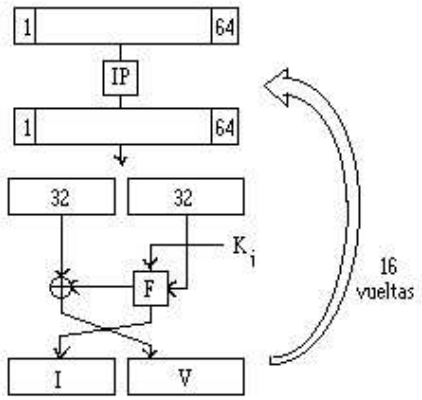
2. Se aplica el algoritmo **DES** propiamente dicho:

   a) Se parte cada bloque en dos (32 bits + 32 bits).

   b) Se aplica un algoritmo tipo Feistel empezando por la derecha 16 veces (16 vueltas/ciclos).

3. Se aplica $\mathbf{P1^{-1}}$.

Figura 1.9: Esquema del funcionamiento del sistema DES.



Una importante cualidad de este algoritmo es que el cifrado y el descifrado *se hacen igual*, esto es, **el propio algoritmo cifra y descifra**.

[11] Con ciertas modificaciones: clave de 56 bits en lugar de 128.

La función de cifrado F(R_i, k_i), en la que se tienen como entradas los 32 bits correspondientes a la mitad derecha de la iteración anterior (R_{i-1}) y los 48 bits de la clave k_i, expande a 48 bits la cadena R_i de 32 bits, permutando y duplicando algunos de ellos para conseguirlo.

La salida de 48 bits de la expansión de R_i y k_i se "suman" en un or exclusivo y la salida se reparte en 8 bloques de 6 bits cada uno. Este resultado sirve como entrada a una *caja S* (en realidad, cada *caja S* contiene 8 cajas cuya entrada son cada uno de los grupos de 6 bits que mencionamos), donde mediante una *transformación no lineal* se obtiene la secuencia de 32 bits (4 bits por cada bloque de 6, que se concatenan a la salida) "final". Es en estas cajas S y en su propiedad de no-linealidad donde reside la potencia de este algoritmo y los que derivan de él.

La respuesta de las cajas S se realiza de acuerdo con matrices o "tablas de conversiones" que cada una tiene asociada. Los 6 bits del bloque que reciben B_j=$b_1b_2b_3b_4b_5b_6$ se transforman en 4, $S_j(B_j)$, correspondientes al valor decimal del elemento de la matriz S_j cuya fila viene determinada por el valor decimal de los bits b_1 y b_6 y cuya columna la define el decimal correspondiente al binario $b_2b_3b_4b_5$ de los bits centrales de B_j.

En cuanto a la función generadora de claves (KS), proporciona una clave k_i=KS(i,K) para cada iteración, donde K es la clave externa de 64 bits, siendo los bits 8, 16, 24,... de paridad impar para protección de dicha clave ante posibles errores de lectura.

Los 64 bits de la clave K se permutan según la permutación PC-1, que sin tomar los bits de paridad permuta los 56 bits restantes, dejando la salida en los registros C_0 y D_0. Los contenidos de estos registros se desplazan a la izquierda una vez para obtener a través de la permutación PC-2 la clave k_1. Las sucesivas claves internas se obtienen a partir de los registros C_i y D_i obtenidos después de realizar los sucesivos desplazamientos a la izquierda de los contenidos de los registros C_{i-1} y D_{i-1}, de forma que:

$$C_i = LS_i(C_{i-1})$$
$$D_i = LS_i(D_{i-1})$$

donde LS_i es un desplazamiento circular a la izquierda en un número de posiciones que está tabulado dependiendo de la iteración de que se trate. La clave interna k_i es, pues, k_i=PC-2($C_i \cdot D_i$).

El descifrado se realiza utilizando el mismo algoritmo descrito, como ya hemos dicho, si bien en la primera iteración se utiliza la clave interna k_{16}, en la segunda la k_{15} y así sucesivamente. Así pues, se invierte el orden de las claves, pero el algoritmo en sí no se invierte.

Como puede intuirse tras su descripción, la única forma de atacar un sistema de cifrado como el DES (y cualquiera de los que se basan en él, como, por ejemplo, IDEA) es mediante *fuerza bruta*, combinada con la utilización de diccionarios, combinaciones de letras, letras y números, etc.

Asimétricos

Los **criptosistemas de clave pública** se caracterizan por utilizar dos claves para cada usuario, una sirve en general para la operación de cifrado y es pública, mientras que la otra clave, la de descifrado, es secreta y es la única que puede recuperar la información cifrada.

Esquema del **RSA** (1978)

Dado un mensaje M...

1. Cada usuario selecciona dos números primos p_u y q_u tales que $n_u = p_u \times q_u$ y $\phi(n) = (p-1)(q-1)$.
2. Se elige la clave pública e_u. Debe cumplir:
   a) $1 \leqslant e_u \leqslant n$
   b) $\text{m.c.d.}(e_u, \phi(n)) = 1$
3. Se aplica el algoritmo extendido de Euclides y se calcula $d_u = \text{m.c.d.}(e_u, \phi(n))$.
4. Se publican n_u $(= p \times q)$ y e_u, y se guardan en secreto d_u, p, q y $\phi(n)$.

Condiciones de selección de p y q

1. La selección de p y q, han de diferir en pocos dígitos.
2. La longitud debe estar en torno a los 200 bits.
3. El $p-1$ y $q-1$ de $\phi(n)$ deben tener factores primos grandes. Eso se garantiza utilizando ***primos seguros***:

$$\begin{gathered} r \text{ primo muy grande} \\ p = 2 \times r + 1 \\ q = 2 \times p + 1 \end{gathered}$$

Bibliografía

[1] Caballero, P. *Introducción a la Criptografía*. RA-MA, 1996.

[2] Lucena López, M. J. *Criptografía y Seguridad en Computadores*. Escuela Politécnica Superior, Universidad de Jaén. 1999.

[3] J.L. Morant Ramón, A. Ribagorda Garnacho, J. Sancho Rodríquez. *Seguridad y Protección de la Información*. Centro de Estudios Ramon Areces, 1994.