UNIVERSIDAD DE MÁLAGA

E.T.S. INGENIERÍA INFORMÁTICA

LENGUAJES Y CIENCIAS DE LA COMPUTACIÓN
UNIVERSIDAD DE MÁLAGA

Tema 8. Introducción a la Criptografía

José A. Montenegro

Dpto. Lenguajes y Ciencias de la Computación
ETSI Informática. Universidad de Málaga
monte@lcc.uma.es twitter

26 de septiembre de 2013

1. Codificación Lenguaje Natural
   - Lenguaje Natural como fuente
   - La incertidumbre del Idioma
   - Redundancia y significado
   - Criptografía 101
   - Análisis de Frecuencias

2. El desarrollo de la criptografía
   - Criptosistemas de Clave Simétrica
   - Cifrado Poli-alfabético
   - El sistema Playfair
   - Algoritmos matemáticos en criptografía
   - Métodos de ataque

3. Criptografía desde el punto teórico y práctico
   - Cifrado en términos de un canal
   - One time pad
   - Métodos Iterativos
   - Estándar de Cifrado
   - El problema de distribución de claves

4. El Criptosistema RSA
   - Viabilidad de RSA
   - Fiabilidad de RSA

Lenguaje Natural como fuente

- Hasta ahora hemos estudiado la codificación para los propósitos de economía y fiabilidad.
- El tercer propósito, envuelve (entre otras cosas) el secreto de los mensajes escritos en un lenguaje natural.
- Para esta razón, las propiedades complejas de un lenguaje juegan una parte importante en la criptografía.
- Comenzamos discutiendo como un lenguaje natural, nos centramos en el Español, puede ser considerado como una fuente en el sentido de la teoría de la codificación.
- Nuestro modelo matemático esta basado en el alfabeto $\mathbb{A}$ denominado *spanish* con 28 símbolos, las letras $A, B, C, \ldots, Z$, y el espacio, denotado por $\sqcup$.
- En el alfabeto utilizado no hacemos distinción entre mayúsculas y minúsculas, y no incluimos signos de puntuación.

Por ejemplo, el texto: Pepe: Juan, vamos a ver el partido!

es traducido a la siguiente cadena en el alfabeto *spanish* .

PEPE␣JUAN␣VAMOS␣A␣VER␣EL␣PARTIDO

Aunque hemos simplificado el alfabeto $\mathbb{A}$, podemos basarnos que *spanish* tiene las propiedades estadísticas que se asemeja al lenguaje Español.

Los valores típicos para estas frecuencia, expresados como el número de ocurrencias por 10000 símbolos, en el lenguaje **Español** son detalladas en la figura 1.

$\sqcup$	A	B	C	D	E	F	G	H
1640	1250	140	460	586	1368	69	101	70
I	J	K	L	M	N	O	P	Q
625	44	10	497	315	671	868	251	88
R	S	T	U	V	W	X	Y	Z
687	798	463	393	90	2	22	90	52

Figura 1 : Tabla de frecuencia de símbolos en Español

- Esta tabla de frecuencias define una distribución de probabilidad p sobre $\mathbb{A}$.
- Pero nos hace falta más información sobre el lenguaje, por ejemplo dada la letra Q, cual es la probabilidad que el próximo símbolo sea U.
- En otras palabras no podemos asumir que los lenguajes son fuentes sin memoria.
- En criptografía, un par de símbolos consecutivos son conocidos como *diagrama* y las frecuencias relativas de los diagramas son significativas.
- Por ejemplo, podemos observar que el diagrama UP sucede más frecuentemente que el diagrama UE. Parte de la tabla de frecuencias para diagramas es mostrado en la figura 2.
- En esta tabla el número en la fila B y columna E es el número de ocurrencias del diagrama BE en una texto de 10.000 símbolos.

	$\sqcup$	A	B	C	D	E	$\ldots$
$\sqcup$	–	212	57	86	69	28	$\ldots$
A	41	–	3	38	22	–	$\ldots$
B	–	3	–	–	–	35	$\ldots$
C	–	41	–	–	–	44	$\ldots$
D	135	22	–	–	–	54	$\ldots$
E	365	57	–	16	73	24	$\ldots$
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

Figura 2 : Parte de Tabla de frecuencia de diagramas en Español

En términos matemáticos, la tabla de frecuencias para los diagramas define una distribución de probabilidad p^2 en $\mathbb{A}^2$.

Podemos observar que para cualquier par ordenados de símbolos *ij*, la probabilidad $p^2(ij)$ no es igual a $p(i)p(j)$.

Por ejemplo, acorde con la tabla anterior,

$$p(A) = 0{,}0724 \text{ y } p(B) = 0{,}0075, \text{ por lo que } p(A)p(B) \approx 0{,}0006,$$

mientras $p^2(AB) \approx 0{,}0003$.

La incertidumbre del Idioma

- Sabemos que una fuente estacionaria tiene la propiedad que cualquier secuencia consecutiva de símbolos tiene la misma probabilidad en todos los puntos del flujo emitido.
- Si concebimos una novela como un flujo de símbolos emitidos por la fuente denominada *spanish*, la propiedad estacionaria nos indica que la probabilidad de encontrar por ejemplo DIA en la página 1 es la misma de encontrarla en la página 99, o en cualquier otra página.
- Formalmente, necesitamos que para cada $n \geq 1$ exista una probabilidad de distribución p^n en el conjunto de n-tuplas $\mathbb{A}^n$.
- Si el flujo emitido por la fuente es representado por la secuencia de variables aleatorias $\xi_1\xi_2\xi_3\ldots$, entonces para cualquier $k \geq 1$ y para cualquier n-tupla de símbolos $x_1x_2\ldots x_n \in \mathbb{A}^n$ necesitamos que

$$Pr(\xi_{k+1} = x_1, \xi_{k+2} = x_2, \ldots, \xi_{k+n} = x_n) = p^n(x_1x_2\ldots x_n).$$

- Es razonable asumir que las propiedades estadísticas del *spanish*, y cualquier otro lenguaje, pueden ser modeladas mediante una representación como una fuente estacionaria, pero sin recoger todas las características del idioma.
- En el caso de un lenguaje natural deberíamos centrarnos (al menos al principio) con codificar un mensaje del alfabeto original por una cadena de símbolos en el mismo alfabeto.
- Por esa razón es apropiado medir la entropía en base a logaritmos en base b, donde b es el número de símbolos en el alfabeto. Por ejemplo, en *spanish* b=28.

Definición 1 (Incertidumbre de un lenguaje natural)

Suponemos que tratamos con un lenguaje natural con un alfabeto de tamaño b como una fuente estacionaria representada por $\mathbf{p}^n (n \geq 1)$. Sea $\mathbb{U}_n = H_b(\mathbf{p}^n)/n$. La Incertidumbre de un lenguaje es definida como

$$\mathbb{U} = inf_{n \in \mathbb{N}} \mathbb{U}_n$$

Nuestra elección de unidades hace que el número este entre 0 y 1. Puede ser visto como la cantidad de incertidumbre por símbolo, cuando la fuente emite un flujo de bloques de tamaño n.

$$\mathbb{U}_n = \frac{H_b(p^n)}{n} = \frac{H_2(p^n)}{n log_2 b}$$

Ejemplo 1

Un lenguaje tiene tres símbolos a,b,c y la tabla de frecuencias para los diagramas es

	a	b	c
a	0,22	0,08	0,10
b	0,10	0,16	0,04
c	0,08	0,06	0,16

Si el lenguajes es considerado como una fuente estacionaria, ¿Cuales son los valores de $\mathbb{U}_1$ *y* $\mathbb{U}_2$*?*

Solución:
$\mathbb{U}_1 = H_3(p^1)$
Para calcular p^1 hacemos uso de los valores de p^2 mediante la suma de la fila correspondiente

$$\begin{aligned} p^1(a) &= 0{,}22 + 0{,}08 + 0{,}10 = 0{,}4, \\ p^1(b) &= 0{,}10 + 0{,}16 + 0{,}04 = 0{,}3, \\ p^1(c) &= 0{,}08 + 0{,}06 + 0{,}16 = 0{,}3. \end{aligned}$$

Por tanto:
$\mathbb{U}_1 = H_3(p^1) = 0{,}4 log_3(1/0{,}4) + 0{,}3 log_3(1/0{,}3) + 0{,}3 log_3(1/0{,}3) \approx 0{,}9912$
Para $\mathbb{U}_2$ tenemos:
$\mathbb{U}_2 = \frac{1}{2} H_3(p^2) =$
$0{,}5\,(0{,}22 log_3(1/0{,}22) + 0{,}08 log_3(1/0{,}08) + \ldots + 0{,}16 log_3(1/0{,}16)) \approx 0{,}9474.$
El hecho que $\mathbb{U}_2 < \mathbb{U}_1$ refleja que la incertidumbre es menor cuando las relaciones entre símbolos consecutivos es tomada en cuenta.

cont. Solución:
Asumiendo que *spanish* es una fuente estacionaria, podemos estimar su incertidumbre utilizando datos experimentales.

Comenzamos con la aproximación de primer orden, que consiste en la distribución p^1 sobre $\mathbb{A}$, tal y como mostramos en la tabla 1.

Este resultado en la estimación

$$\mathfrak{U}_1 = H_{28}(p^1) = \sum_{i \in \mathbb{A}} (p^1)(i) log_{28}(1/p^1(i)),$$

el cual tiene una aproximación sobre 0.85.

La aproximación de segundo orden es determinada por la tabla de frecuencias para los diagramas, las cuales definen una distribución p^2 en el conjunto $\mathbb{A}^2$, sería dado por

$$\mathfrak{U}_2 = \frac{1}{2} H_{28}(p^2) = \sum_{ij \in \mathbb{A}^2} (p^2)(ij) log_{28}(1/p^2(ij)),$$

cont. Solución:
Después de realizar los cálculos, obtenemos un valor aproximado de 0.70.

Desafortunadamente, realizar estimaciones fiables $\mathfrak{U}_n$ para valores grandes de n es muy difícil, debido a que hay 28^n contribuciones individuales a la entropía, y la mayoría de ellos son muy pequeños.

Sin embargo es razonable esperar que si secuencias más largas son tomadas en cuenta, la estimación de incertidumbre por símbolo se decrementará.

Solo a modo de ejemplo, y tras muchos cálculos obtenemos

$$\mathfrak{U} = inf_{n \in N} \mathfrak{U}_n \approx 0{,}3$$

que nos indica que, si nos proporcionan un trozo de texto, entonces podemos predecir que viene detrás con una certeza cercana al $70\,\%$.

Redundancia y significado

- Hay buenas razones para creer que la propiedad de ser una fuente estacionaria no captura todas las características de un lenguaje natural.
- Además de sus propiedades estadísticas no triviales, un lenguaje natural tiene otra propiedad fundamental, la cual podemos referirnos como *significado*.
- Aunque está claro que el objetivo de un lenguaje natural es transportar mensajes con significado, no es fácil expresar esta idea en términos matemáticos.
- Una dificultad es que es bastante posible generar cadenas de símbolos que no tienen sentido, aunque tengan propiedades estadísticas similares a la fuente que hemos denominado *spanish*.
- Supongamos que tomamos un libro, asumiendo que es una típica salida de *spanish*. Si reordenamos las páginas del libro, y las oraciones de cada página, entonces tendremos una salida que es virtualmente indistingible (desde el punto de vista estadístico) al libro original. Pero claramente, no tendrá sentido.

- Una propiedad característica de un mensaje con sentido es que es posible acortarlo sin perder el significado. En un nivel muy básico esta es la razón por que utilizamos abreviaciones, tales como EEUU. Similarmente, omitir algunas letras de un mensaje con significado no previene de que sea entendido.
- Esta propiedad de un lenguaje natural es conocida como *redundancia*.
- Ahora proporcionaremos un argumento heuristico, aportado por Shannon, que vincula la redundancia de un lenguaje con su incertidumbre.
- Consideremos el conjunto de mensajes con significado de longitud n emitido por la fuente que denominamos *spanish*.
- Supongamos que ha sido encontrado por un experimento que la proporción de símbolos que pueden ser emitidos de estos mensajes sin destrozar el significado es f_n.

- El proceso de eliminar nf_n símbolos puede ser pensado como la codificación del mensaje de longitud n por un mensaje en el mismo alfabeto, pero con longitud

$$l_n = n(1 - f_n).$$

Ahora el teorema para fuentes estacionarias (Teorema 4.3) dice que el valor optimo de l_n/n es cercano a U, la incertidumbre de la fuente. Es por tanto razonable asumir que

$$inf_{n \in N} f_n = inf_{n \in N}(1 - l_n/n) = 1 - U.$$

Definición 2 (Redundancia)

La redundancia de un lenguaje natural con b símbolos está definida como

$$R = 1 - U$$

donde U es la incertidumbre.

- Hay varias formas de mirar a esta definición. Una de ellas es relativa a una forma alternativa de calcular U, utilizando los resultados experimentales en la redundancia.
- La mayoría de los experimentos sugieren que sobre la mitad de las letras en un mensaje largo pueden ser omitidas, aunque los resultados pueden varia acorde con la regla que es utilizada para suprimir las letras.
- Esta observación es consistente con la estimación $U \approx 0{,}3$ obtenida anteriormente.

- A modo de resumen, hay varias formas de estimar la incertidumbre y redundancia de un lenguaje natural como el español.
- Sin embargo, ni los lingüistas teóricos ni los matemáticos han conseguido con éxito en formular una teoría que capture todas las sutilezas de un lenguaje natural, pro lo que las estimaciones numéricas son muy imprecisas.
- Más importante es el hecho que los conceptos de redundancia y significado son muy relevantes para los aspectos prácticos de la criptografía.

Criptografía 101

- Ahora estamos listos para discutir los aspectos tradicionales de la criptografía: el estudio de la codificación para el propósito de transmitir mensajes secretos escritos en un lenguaje natural.
- Posiblemente por su ya establecida historia, la criptografía ha desarrollado su propia terminología.
- Comenzaremos con una breve descripción de algunos métodos que han sido usados en el pasado, y así podremos definir la mayoría de la terminología.
- Uno de los sistemas criptográficos más antiguos que tenemos referencia fue utilizado por Julio Cesar hace casi 2000 años.

- Para un mensaje emitido por la fuente que hemos denominado *spanish*.
- El emisor elige un número k y reemplaza cada letra por una que está k lugares posteriores, en orden alfabético.
- Para simplificar la explicación asumiremos que el espacio ␣ y ñ no son modificadas.
- Por ejemplo, cuando k=5 las letras son modificadas como sigue:

␣ 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
␣ 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

- Utilizando esta regla, el mensaje

MALAGA ␣ CAMPEON es RFQFLF ␣ HFRUJTS.

- En criptografía este proceso es conocido como *cifrado*, y el número k es la clave.
- En términos matemáticos, el cifrado es simplemente codificación, con la característica adicional que la función de codificación E_k depende del parámetro k.
- Por ejemplo, en nuestra versión del sistema Cesar E_k es la función $\mathbb{A} \to \mathbb{A}$ definida mediante

$$E_k(x) = [x + k] \ (x \neq \sqcup) \ , \ E_k(\sqcup) = \sqcup,$$

donde [x + k] representa el símbolo que está k posiciones de x. La función E_k es extendida para que actúe sobre los mensajes (cadenas de símbolos) mediante concatenación:

$$E_k(DUDA) = E_k(D)E_k(U)E_k(D)E_k(A).$$

Necesitamos que la función extendida E_k es una función inyectiva, por lo que la codificación es unívocamente decodificable. La situación general es como sigue:

Definición 3 (Funciones Cifrado)

Sea $\mathcal{M}$ y $\mathcal{C}$ conjuntos de mensajes y sea $\mathcal{K}$ un conjunto.

Supongase que para cada $k \in \mathcal{K}$ hay una función inyectiva tal que $E_k : \mathcal{M} \to \mathcal{C}$.

Entonces diremos que $\{E_k\}$ es un conjunto de funciones de cifrados, y un elemento $k \in \mathcal{K}$ es denominada una clave.

En el sistema de Cesar $\mathcal{M}$ y $\mathcal{C}$ son tomados del conjunto $\mathbb{A}^*$ de cadenas de símbolos en $\mathbb{A}$, y $\mathcal{K}$ es el conjunto de símbolos en el rango $1 \leq k \leq 25$.

- La principal lección aprendida en la historia de la criptografía es que es una mala idea ocultar el método de cifrado, la instacia de las funciones E_k.
- Si un número de personas quieren comunicarse de forma segura, no pueden basarse en ocultar el sistema que están usando.
- La seguridad del sistema debe depender solamente del valor específico que han acordado asignar a la clave k.

- Actualmente las partes involucradas en la comunicación reciben los nombres de Alice, Bob y Even.
- Básicamente, Alice es el emisor del mensaje, Bob es el receptor, y Eve es un adversario que intenta intervenir de alguna forma en la comunicación.
- Dentro de este marco de trabajo podemos distinguir varios requisitos de seguridad.
- El requisito más obvio es la *Confidencialidad*: un mensaje de Alice a Bob no debe ser entendido por Eve. Este es el requisito que vamos a considerar inicialmente.
- Sin embargo, la criptografía moderna también cubre otros aspectos de la seguridad, como Autenticidad, Integridad y No repudio.

- Vamos asumir que Alice y Bob están de acuerdo con un método de cifrado y el valor de la clave k, por lo tanto ellos conocen la función de cifrado E_k.
- El mensaje que Alice quiere mandar a Bob es conocido como *texto en claro (plaintext)*, y por ahora podemos asumir que está expresado en un lenguaje natural, tal como el español.
- Alice utiliza la función E_k para transformar el texto en claro en una forma codificada, conocido como *texto cifrado (ciphertext)*, y lo envía a Bob.
- Por lo que si Alice y Bob están de acuerdo en utilizar el sistema Cesar con k=5, y el texto en claro es

$$\text{MALAGA} \sqcup \text{CAMPEON}$$

entonces Bob recibirá el texto cifrado

$$\text{RFQFLF} \sqcup \text{HFRUJTS}.$$

En este caso es fácil como Bob puede obtener su mensaje en claro, pero es útil considerar la situación en términos más generales.

- Suponemos que tenemos un conjunto de funciones de cifrado $E_k : \mathcal{M} \to \mathcal{C}$.
- Tal y como hemos establecido anteriormente cada E_k es una función inyectiva.
- Esto significa que si $E_k(m) = c$, entonces m es el único texto en claro que E_k cifra como c.
- Consecuentemente, existe una función F, definida como imagen de E_k, que toma c a m:

$$F(E_k(m)) = m \text{ para todo } m \in \mathcal{M}.$$

- Formalmente, F es la inversa de E_k. En la práctica no es suficiente con conocer que existe una inversa, es necesario tener una regla explícita para calcularla.

Definición 4 (Funciones Descifrado)

Sea $\{E_k\}$ $(k \in \mathcal{K})$ *sea un conjunto de funciones de cifrado* $\mathcal{M}$ *y* $\mathcal{C}$.

Entonces el conjunto de funciones $\{D_l\}$ $(l \in \mathcal{K})$ *es un conjunto de funciones de descifrado para* $\{E_k\}$ *si para cada* $k \in \mathcal{K}$ *existe un* $l \in \mathcal{K}$ *tal que* D_l *es la inversa de* E_k*:*

$$D_l(E_k(m)) = m \text{ para todo } m \in \mathcal{M}$$

En el sistema Cesar, tomaremos l tal que $l = -k(mod28)$ y la inversa de E_k es

$$D_l(x) = [x + l](x \neq \sqcup),\ D_l(\sqcup) = \sqcup.$$

- Ya que Bob conoce k, el también conoce l, y puede descifrar el mensaje de Alice.
- Notese que en este caso la función de descifrado D_l y la función de cifrado E_k tienen la misma forma, aunque no tiene que ser así de forma general.
- Ahora consideramos la situación cuando Eve intercepta el mensaje cifrado

$$\text{RFQFLF} \sqcup \text{HFRUJTS}.$$

- Ella no puede simplemente aplicar una función de descifrado y recuperar el texto en claro, ya que ella no conoce cual clave ha sido usada.
- Si ella quiere obtener el texto en claro, el método más fácil sería intentar encontrar una clave.

- El proceso para obtener el correspondiente texto en claro para algún texto cifrado, es encontrar el valor de la clave k o algún método indirecto, para romper el sistema.
- Cualquier método el cual Eve pueda llevar a cabo es conocido como un ataque.
- Hemos observado que Eve puede conocer que sistema se está utilizando, por tanto, si ella conoce la clave, entonces ella sabrá como usarla.
- Para el sistema Cesar existe un ataque simple por el método conocido como búsqueda exhaustiva.
- Debido a que solamente los posibles valores de k son $1, 2, 3 \ldots, 28$, es fácil encontrar cual es el correcto.
- Asumiendo que el texto en claro es un mensaje expresado en un lenguaje natural, Eve conocerá cuando la clave correcta ha sido encontrada, ya que el mensaje tendrá sentido.

Ejemplo 2

Suponemos que Eve intercepta el mensaje cifrado

$$GJOBOTG \sqcup WAK \sqcup VUTK$$

¿Cual es el correspondiente mensaje en claro?

Solución:
Cuando utilizamos la clave k=1,2,3, . . . en la parte inicial del texto cifrado, el resultado no tiene significado hasta que k=6 es alcanzado.

$$k = 1\ FINANSF \sqcup \ldots$$
$$k = 2\ EHMZMRE \sqcup \ldots$$
$$k = 3\ DGLYLQD \sqcup \ldots$$
$$k = 4\ CFKXKPC \sqcup \ldots$$
$$k = 5\ BEJWJOB \sqcup \ldots$$
$$k = 6\ ADIVINA \sqcup \ldots$$

Por lo que intentando con la clave k=6 en el resto del mensaje, este produce una mensaje con significado y el texto en claro es

$$ADIVINA \sqcup QUE \sqcup PONE$$

Análisis de Frecuencias

- ¿Como es posible que Alice y Bob se defiendan del ataque mediante una búsqueda exhaustiva?
- Una posible idea es incrementar el número de claves.
- El sistema Cesar utiliza la regla $x \mapsto [x + k]$ para reemplazar las letras, y hay solamente 28 valores posibles de la clave k.
- Pero claramente cada permutación de 26 letras puede ser usada. Por lo que tenemos 28! permutaciones,

  por tanto Eve necesitará unos recursos sustanciales si quiere probar con todos ellos.

En criptografía un sistema que solamente permuta las letras del alfabeto en una forma determinada es conocido como sustitución *mono-alfabetica*.

La clave es una permutación σ de 28 letras y la función de cifrado puede ser definida como

$$E_\sigma(x) = \sigma(x) \ (x \neq \sqcup), \ E_\sigma(\sqcup) = \sqcup$$

La inversa de E_σ es la función D_τ que utilizar la permutación inversa $\tau = \sigma^{-1}$:

$$D_\tau = \tau(x) \ (x \neq \sqcup), \ D_\tau(\sqcup) = \sqcup$$

- Es conveniente utilizar una clave que puede ser memorizada, y esto es realizado escogiendo una palabra clave. Por ejemplo, si la palabra clave escogida es CAMPEON, la permutación correspondiente σ es definida como sigue:

$$
\begin{array}{l}
\sqcup\ \mathrm{A\ B\ C\ D\ E\ F\ G\ H\ I\ J\ K\ L\ M\ N\ O\ P\ Q\ R\ S\ T\ U\ V\ W\ X\ Y\ Z} \\
\sqcup\ \mathrm{C\ A\ M\ P\ E\ O\ N\ B\ D\ F\ G\ H\ I\ J\ K\ L\ Q\ R\ S\ T\ U\ V\ W\ X\ Y\ Z}
\end{array}
$$

- Aunque el ataque por búsqueda exhaustiva necesita de muchos recursos, otro método de ataque, conocido como análisis de frecuencias, es usualmente más efectivo. Consta que este método fue utilizado por criptógrafos Arabes en el siglo 9.
- El método esta basado en el hecho que un mensaje en claro de un lenguaje natural tiene propiedades estadísticas. Además, es razonable asumir que el texto transporta un mensaje con sentido.
- En cualquier mensaje de longitud razonable, los símbolos, diagramas y demás, suceden con frecuencias cercanas a aquellas dadas en las tablas estándar de frecuencias.
- Por tanto si el símbolo x sucede con una frecuencia n_x en el texto en claro, y la función de cifrado es E_σ, el símbolo $\sigma(x)$ ocurrirá con una frecuencia n_x en el texto cifrado.
- Utilizando los datos proporcionados por esta observación, y el hecho que el texto en claro tiene algún sentido, el texto cifrado puede ser descifrado.

En resumen, el ataque por análisis de frecuencia en una pieza de un texto cifrado producido por una sustitución monoalfabetica es realizado como sigue:

- Contar las frecuencias de los símbolos en el texto cifrado;
- Comparar estas con las frecuencias estándar dadas en las tablas;
- Verificar la correspondencia de los símbolos, hasta que un texto con sentido es obtenido

El tercer paso involucra realizar y verificar una hipótesis, y no existe una regla definitiva para que eso pueda ser realizado.

Criptosistemas de Clave Simétrica

- Anteriormente hemos descrito un marco de trabajo para la criptografía basado en un conjunto de mensajes en claro $\mathcal{M}$, un conjunto de mensajes cifrados $\mathcal{C}$ y un conjunto de claves $\mathcal{K}$.
- Para cada $k \in \mathcal{K}$ hay una función de cifrado $E_k : \mathcal{M} \to \mathcal{C}$, con su correspondiente inversa la función de descifrado D_l. En otras palabras

$$D_l(E_k(m)) = m \text{ para toda } m \in \mathcal{M}$$

- Nos referimos a este marco de trabajo como criptosistema.
- Los criptosistemas pueden ser construidos e implementados de muchas formas.

- En los sistemas de sustitución mono alfabéticos discutidos anteriormente, las claves correspondientes k y l son permutaciones inversas, y por tanto están vinculadas de una forma muy simple. Durante, muchos siglos se ha asumido que cualquier criptosistema práctico debe tener una propiedad similar.
- Ahora está claro que esa presunción es injustificada, pero sistemas con esta propiedad son todavía ampliamente utilizados, y le daremos a ellos un nombre.

Definición 5 (Criptosistema Clave Simétrica)

Supongamos que tenemos un criptosistema en el cual D_l es la inversa de E_k.

El sistema es denominado criptosistema de clave simétrica si, cualquiera de las dos k o l es conocida, entonces es fácil de conocer la otra.

- Durante muchos años los criptógrafos trabajaron para hacer que los sistemas sean seguros ante un ataque de análisis.
- En un sistema mono alfabético, con una clave dada, la letra E (por ejemplo) siempre será reemplazada por la misma letra en el texto cifrado, por ejemplo X.
- El principal problema que presenta es que X será el carácter que más aparece debido a que ha reemplazado a una de las letras que más aparecía en el texto en claro.
- Esta debilidad del sistema puede eliminarse si utilizamos un sistema en el cual E no siempre reemplazado por la misma letra.

Definición 6 (Cifrado Poli alfabetico)

En un sistema poli alfabetico la regla para el cifrado es que cada letra en el texto en claro es reemplazado por una letra que depende, no solamente por la letra en sí mismo, también por su posición en el texto.

- Un método simple de cifrado poli alfabetico fue desarrollado en el siglo 16, y es conocido como el *Sistema Vigenere*.
- Básicamente utiliza las permutaciones realizadas en el Sistema Cesar, que son desplazamiento cíclicos de tamaño $(1 \leq k \leq 28)$, denotado en el capítulo anterior como $E_k(x) = [x + k]$.
- Como en los casos anteriores, eliminaremos los espacios, resultando un texto en claro con todas las letras juntas.

- En el *Sistema Vigenere* la clave es una secuencia $K = (k_1, k_2, \ldots, k_m)$ de números. Si el texto en claro es $x_1 x_2 x_3 \ldots x_i \ldots$, la regla para el cifrado es

$$E_K(x_i) = [x_i + k_j] \text{ si } i \equiv j \ (mod\, m).$$

- Por ejemplo, si la clave tiene longitud m=4, entonces $x_1, x_5, x_9 \ldots$ son cifrados con un desplazamiento de k_1; $x_2, x_6, x_1 0 \ldots$ son cifrados con un desplazamiento de k_2 y así sucesivamente.
- Esta descripción representa un sistema de clave simétrica, ya que el descifrado es realizado mediante el reverso de los desplazamientos.
- En la práctica la clave K es usualmente expresado como una palabra, donde las letras de la palabra representan los desplazamientos relevantes.

Ejemplo 3

Tenemos la palabra clave $CLAVE$, que representa la clave

$$K = \{3, 12, 1, 22, 5\}$$

¿Cual será el cifrado del siguiente texto?

$$VOTARCAMBIAALGO$$

Solución:

V	O	T	A	R	C	A	M	B	I	A	A	L	G	O
3	12	1	22	5	3	12	1	22	5	3	12	1	22	5
Y	A	U	W	W	F	M	N	X	N	D	M	M	C	T

- Durante varios siglos el sistema Vigenere fue concebido como irrompible, y fue denominado como el *cifrado indescifrable*.
- Sin embargo, el sistema posee una vulnerabilidad inherente, debido a que la elección de los desplazamientos deben de estar basados en una regla definitiva, ya que de otra forma el receptor no será capaz de descifrarlo.
- Esta debilidad puede ser explotada para montar un ataque.

- Aunque se pensaba que era irrompible, el Sistema Vigenere era difícil de utilizar. Por ese motivo existía un considerable interés en métodos alternativos.
- Uno de esos métodos era el sistema Playfair, en el cual los diagramas son permutados, en vez de las letras individuales.
- Una clave para el sistema es derivada de una matriz 5x5 que contiene 25 letras (Para tener 25 letras, podemos identificar la Y como W).
- Si no permitimos diagramas con letras dobles de la forma LL, el número de diagramas posibles es 25 x 24 =600, dando la posibilidad de unas 600! claves, por lo cual una búsqueda exhaustiva de la clave es una tarea complicada.

- El orden de las letras en la matriz puede ser aleatorio, o puede ser definida por una palabra clave.
- Figura 3 muestra la matriz basada en la palabra clave $PERSONAL$.
- De esta forma tanto receptor como emisor pueden recordar la clave, y siguiendo unas simples reglas, los procedimientos de cifrado y descifrado son directos.

$$\begin{array}{ccccc} P & E & R & S & O \\ N & A & L & B & C \\ D & F & G & H & I \\ J & K & M & Q & T \\ U & V & W & X & Z \end{array}$$

Figura 3 : La matriz Playfair con la palabra $PERSONAL$

En el proceso de cifrado los espacios son ignorados, y las letras dobles son separadas mediante una letra comodín tal como X o Z. El texto en claro es dividido en diagramas, cada diagrama xy es cifrado acorde con las siguientes reglas.

Caso 1 Suponemos que x e y están en distintas filas y distintas columnas. Entonces $xy \mapsto ab$ donde x,y,a,b son las esquinas de un rectángulo, y x,a están en la misma fila.

Caso 2 Suponemos que x,y están en la misma fila. Entonces $xy \mapsto uv$, donde u y v son las letras que están a la derecha de x e y respectivamente. (Si alguno de x o y están al final de la fila, la letra al inicio de la fila es la elegida).

Caso 3 Suponemos x, y están en la misma columna. Entonces $xy \mapsto uv$, donde u y v son las letras inmediatamente debajo de x e y respectivamente. (Si alguno de los dos (x o y) están al final de la columna, la letra al inicio de la columna es utilizada.

- Por ejemplo,

$$QA \mapsto KB,\ FW \mapsto GV,\ DI \mapsto FD,\ SH \mapsto BQ$$

- El sistema Playfair es un sistema de clave simétrica.
- La clave de cifrado y descifrado son permutaciones del conjunto de 600 diagramas posibles, Alice utiliza las reglas dadas anteriormente para construir la clave para cifrado y la matriz dada, y Bob utiliza la misma matriz, pero con reglas ligeramente diferentes para construir las claves para el descifrado.
- Las dos claves son permutaciones inversas.

Ejemplo 4

Supongamos que recibimos el siguiente texto cifrado, conociendo que el sistema Playfair ha sido utilizado y la palabra clave es PERSONAL. ¿Cual era el texto en claro?

$$DCKOSREBJCARLNBLORISXZ$$

Solución:
El texto cifrado es dividido en una secuencia de diagramas como:

DC KO SR EB JC AR LN BL OR IS XZ

Utilizando las reglas de forma inversa, el correspondiente secuencia de diagramas es

IN TE RE SA NT EL AC LA SE HO YX

Donde podemos intuir que el texto es:

INTERESENTE LA CLASE HOY

- Como podemos intuir, es posible atacar y romper el sistema $Playfair$ mediante análisis de frecuencias, utilizando las tablas de frecuencias para diagramas y siguiendo el mismo procedimiento que detallamos anteriormente.
- Sin embargo, este procedimiento requiere más datos que en el caso de análisis de un simple símbolo, y consumirá más tiempo completarlo.
- Por tanto, el sistema puede proporcionar un nivel aceptable de seguridad en determinadas circunstancias.

Algoritmos matemáticos en criptografía

- Hasta ahora hemos usado las matemáticas principalmente con el propósito de explicar como funciona ciertos criptosistemas. En la criptografía actual, las matemáticas tiene un uso más fundamental.
- Anteriormente, en el problema de crear códigos con buenas propiedades de corrección de errores, observamos que los símbolos 0 y 1 en el alfabeto binario $\mathbb{B}$ tenían propiedades algebraicas.
- En tal caso introducimos la notación $\mathbb{F}_2$ para el cuerpo cuyos elementos son 0 y 1 y $\mathbb{F}_2^n$ para el espacio de vectores de n-tuplas sobre $\mathbb{F}_2$.
- La utilización de estas construcciones algebraicas nos permiten extender el rango de métodos que podemos emplear.

- En general, los símbolos en un mensaje pueden ser representados por los elementos de una estructura algebraica de muchas formas.
- La forma más simple de representar un conjunto de n objetos mediante una estructura algebraica es utilizar enteros modulo n, denotado como $\mathbb{Z}_n$.
- Por ejemplo, el alfabeto $\mathbb{A}$ puede ser representado mediante $0, 1, 2, \ldots, 26$, denotados como elementos de $\mathbb{Z}_{27}$.
- Los enteros $mod\ n$ pueden ser sumados y multiplicados de forma que ellos satisfacen las reglas de la aritmética. Técnicamente, decimos que $\mathbb{Z}_n$ es un *anillo*.

- En realidad, a menudo es conveniente utilizar un alfabeto con un *número primo de símbolos*, ya que cuando p es un número primo, los enteros $mod\ p$ forman un cuerpo.
- Esto significa cada elemento que no es cero tiene un inverso en la operación de multiplicación. Para enfatizar su naturaleza especial denotaremos este cuerpo como $\mathbb{F}_p$.
- Por tanto si extendemos el alfabeto $\mathbb{A}$, permitiendo mensajes que contienen comas y puntos, además de los usuales 27 símbolos, entonces tendremos 29 símbolos, un número primo.
- Podemos representar el espacio $\sqcup$ por el 0, las letras $A, B, \ldots, Z$ por $1, 2, \ldots, 26$, y la coma y el punto por 27 y 28, todos considerados como elementos del cuerpo $\mathbb{F}_{29}$.

- Por ejemplo, si el mensaje es

$$I \sqcup CAME, \sqcup I \sqcup SAW, \sqcup I \sqcup CONQUERED.$$

lo podemos reemplazar por:

9 0 3 1 13 5 27 0 9 0 13 1 23 27 0 9 0 3 15 14 17 21 5 18 5 4 28

- El punto es que, aunque la aritmética no tiene significado cuando los símbolos son 'únicamente'símbolos, podemos realizar todas las operaciones aritméticas, incluyendo una división por un número distinto de cero. Esta situación incrementa enormemente el rango de métodos de cifrado que están disponibles.
- Otra posibilidad es cuando utilizamos la técnica estándar de dividir el flujo de símbolos en bloques con un tamaño apropiado m, por lo que cada bloque será un m-vector sobre $\mathbb{F}_{29}$. Ahora, no solo es posible aplicar la aritmética en símbolos individuales, podemos utilizar algebra lineal para manipular los vectores.

Ejemplo 5

Representamos el mensaje $I \sqcup CAME, \sqcup I \sqcup SAW, \sqcup I \sqcup CONQUERED$, *como una cadena de 2-vectores en* $\mathbb{F}_{29}$, *y establecemos las cadenas de 2-vectores que resultan cuando aplicamos la matriz K. ¿Como podemos transformar al mensaje original el vector resultante?.*

$$K = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix}$$

Solución:
El mensaje original es representado por:

$$[9\,0]'[3\,1]'[13\,5]'\ldots$$

Y mediante las operaciones

$$K = \begin{pmatrix} 3 & 4 \\ 2 & 3 \end{pmatrix} \begin{pmatrix} 9 \\ 0 \end{pmatrix} = \begin{pmatrix} 27 \\ 18 \end{pmatrix}$$

son convertidos en

$$[27\,18]'[13\,9]'[1\,12]'\ldots$$

Dado la anterior cadena, podemos recuperar la primera aplicando la matriz inversa, la cual en este caso es

$$K^{-1} = \begin{pmatrix} 3 & 25 \\ 27 & 3 \end{pmatrix}$$

- El ejemplo ilustra la básica idea que hay detrás de un acercamiento inicial para utilizar técnicas de matemáticas abstractas en criptografía.
- En dos trabajos publicados sobre 1930, *L.S. Hill* sugería que los mensajes podrían ser cifrados mediante la aplicación de una transformación lineal, es decir, multiplicando los vectores por un matriz.
- Generalmente, en el sistema de Hill la clave es una matriz K invertible $m\ x\ n$ y la función de cifrado es

$$E_K(x) = K_x,$$

  donde x es un m-vector.
- Si $y = Kx$, entonces $K^{-1}y = x$. Por tanto, la función de descifrado correspondiente a E_K es dada por $D_L(y) = Ly$, donde $L = K^{-1}$.
- El sistema de Hill es otro ejemplo de sistema de clave simétrica, ya que K^{-1} puede ser calculado desde K utilizando los métodos estándar de la algebra matricial.

Métodos de ataque

¿Como puede ser atacado el sistema Hill?

- Un ataque por búsqueda exhaustiva, requiere verificar todas las matrices invertibles m x m sobre $\mathbb{F}_{29}$.
- Una matriz m x m tiene m^2 componentes, y si cada componente es un elemento de $\mathbb{F}_{29}$, el número de posibilidades es 29^{m^2}.
- La mayoría de todas estas matrices son invertibles, y consecuentemente m puede ser escogido de forma que un ataque por búsqueda exhaustiva no es no posible.

- Si la búsqueda exhaustiva no es posible, un posible ataque a un criptosistema de clave simétrica es conocido como ataque de texto cifrado.
- Eve obtiene una pieza de texto cifrado c e intenta utilizar esta información para encontrar las claves de cifrado k y descifrado l.
- Si encontramos estas claves, Eve no solamente puede descifrar el texto cifrado conocido, utilizando la regla $D_l(c) = m$, también otros textos cifrados que utilizan las mismas claves.

- El ataque mediante análisis de frecuencias es conocido como ataque de texto cifrado conocido. Pero en el sistema Hill este ataque no es posible ya que las letras han sido mezcladas.
- Dada cualquier secuencia de letras puede ser cifrado de muchas formas distintas, dependiendo de la posición en un bloque y las otras letras que suceden en ese bloque.
- Sin embargo, un punto débil del sistema Hill es la posibilidad de otro ataque: puede ser posible para Eve obtener algunos trozos del texto en claro m y los correspondientes texto cifrado c.

Los criptógrafos reconocen dos tipos de ataques basados en esta información:

1. En un ataque por texto en claro conocido, Eve ha obtenido algún par (m_i, c_i), donde cada m_i es un texto en claro y c_i es el correspondiente texto cifrado.
2. En un ataque de texto en claro elegido, Eve ha obtenido los textos cifrados c_i correspondiente a un número específico de textos en claro m_i que ella ha elegido.

Ejemplo 6

Supongase que Eve intercepta algún texto cifrado el cual contiene un informe sobre una emboscada, utilizando el sistema Hill con m=2. Eve sospecha que el texto cifrado

$$[1515]'[2515]'[173]'$$

representa la palabra $AMBUSH$ (emboscada). ¿Como puede probar su hipótesis y encontrar la clave?

Solución:
En 2-vectores sobre $\mathbb{F}_{29}$ la palabra $AMBUSH$ es representada mediante

$$[113]'[221]'[198]'.$$

Sabemos que la clave es

$$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

La hipótesis de Eve es que la secuencia $[113]'[221]'[198]'$ es cifrada como $[1515]'[2515]'[173]'$. Si la hipótesis es verdadera, los bloques AM y BU son cifradas como:

$$\begin{pmatrix} 15 \\ 15 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 \\ 13 \end{pmatrix}, \begin{pmatrix} 25 \\ 15 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2 \\ 21 \end{pmatrix}$$

cont. Solución:

Consecuentemente

$$\begin{pmatrix} 15 & 25 \\ 15 & 15 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 13 & 21 \end{pmatrix},$$

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 15 & 25 \\ 15 & 15 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 13 & 21 \end{pmatrix}^{-1}.$$

Eve haciendo cálculos de algebra elemental en $\mathbb{F}_{29}$, puede calcular

$$K = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix}.$$

cont. Solución:

Eve puede verificar su hipótesis verificando que, con esta K, el tercer bloque SH es cifrado como esperaba:

$$\begin{pmatrix} 2 & 1 \\ 5 & 3 \end{pmatrix} \begin{pmatrix} 19 \\ 8 \end{pmatrix} = \begin{pmatrix} 17 \\ 3 \end{pmatrix}$$

Ya que ha funcionado entonces es posible aplicar,

$$K^{-1} = \begin{pmatrix} 3 & -1 \\ -5 & 2 \end{pmatrix}$$

al resto del texto cifrado. Si obtenemos un texto en claro con significado, el problema está resuelto. En caso contrario, Eve puede intentar con un trozo de texto distinto.

Cifrado en términos de un canal

- Los continuos fallos para construir un sistema que no se pueda romper lleva al análisis matemáticos de los procesos de cifrado.
- Shannon propone un marco de trabajo basado en la idea que convertir texto en claro en texto cifrado puede ser representado como el envío de la información por un canal.
- Asumimos que hay un conjunto finito M de mensajes en claros que son elegibles para ser enviados.
- Denotamos p_m la probabilidad que el mensaje es m, en otras palabras, hay una distribución de probabilidad p en M.
- Denotaremos como (M, p) una fuente sin memoria, el cual es la entrad a un canal, como explicamos a continuación.

- Sea el conjunto de todas las posibles claves K, y sea E_k la función de cifrado para la clave $k \in K$.
- Supongamos que la probabilidad que E_k sea utilizada es r_k, por lo que tenemos una distribución de probabilidad r sobre el conjunto K.
- Asumimos que la elección de k es independiente del mensaje m.
- La salida del sistema es el texto cifrado $c = E_k(m)$, y por tanto la probabilidad que c ocurra es

$$q_c = \text{Pr(texto cifrado es c)} = \sum p_m r_k,$$

  donde la suma es definida sobre el conjunto de pares (m,k) tal que $E_k(m) = c$.
- Por tanto tenemos una distribución q en el conjunto C, y podemos establecer como (C,q) como la salida de un canal (Figura 4).

Figura 4 : Cifrado representado como transmisión a través de un Canal

- Si establecemos p y q como vectores de la forma usual como llevamos utilizando, la transformación realizada por el proceso de cifrado es definida por una matriz Γ tal que $q = p\Gamma$.
- Comparando con la fórmula para q_c dada anteriormente muestra que podemos definir Γ como sigue:

$$\Gamma_{mc} = Pr(c|m) = \sum_k r_k$$

  donde la suma es realizada sobre todas las claves k tales que $E_k(m) = c$.
- Notese que Γ depende de la distribución r, de la misma forma que la matriz del canal para el BSC extendido depende de la probabilidad de error e.

Ejemplo 7

Sea $M = C$ el vector del espacio $\mathbb{F}_2^2$ de pares ordenados xy, y tomaremos K=$\{1,2\}$ con $r_1 = r$, $r_2 = 1 - r$. Supongamos que las funciones de cifrado son $E_1(xy) = xy + 01$, $E_2(xy) = xy + 10$,
Si la distribución de entrada p en $\mathbb{F}_2^2$ es dada por

00	*01*	*10*	*11*
0.1	*0.2*	*0.3*	*0.4*

¿Cual es la distribución de probabilidad q?

Solución

Notese que en este ejemplo hay al menos un k para cada par (m.c) y ordenando las filas y columnas en el orden 00,01,10,11 la matriz del canal Γ es

$$\begin{pmatrix} 0 & r & 1-r & 0 \\ r & 0 & 0 & 1-r \\ 1-r & 0 & 0 & r \\ 0 & 1-r & r & 0 \end{pmatrix}$$

Por tanto la distribución de salida $q = p\Gamma$ es

00	01	10	11
0.3-0.1r	0.4-0.3r	0.1+0.3r	0.2+0.1r

- El método básico de atacar a un criptosistema es el ataque del texto cifrado conocido, en el cual Eve obtiene un trozo del texto cifrado c e intenta encontrar la clave k.
- Por tanto consideramos la incertidumbre sobre k dado c, o en términos usados anteriormente, la entropía condicional $H(r|q)$, que en este contexto tiene un nombre especial.

Definición 7 (Equivocación Clave)

Dada las distribuciones de probabilidad p en M y r en K, la equivocación en la clave del sistema representada por Γ *es* $H(r|q)$*, donde* $q = p\Gamma$.

Lema 1

En la anotación anterior, la equivocación en la clave es dada por $H(r|q) = H(r) + H(p) - H(q)$.

Secreto Perfecto

- Es razonable decir que un secreto perfecto ocurre cuando la incertidumbre sobre un trozo de texto no es alterado si el correspondiente texto cifrado es conocido.
- En el formalismo del canal, esto es traducido al hecho que si la incertidumbre asociada con la fuente de entrada (M,p) debería ser la misma que su incertidumbre cuando la fuente de salida (C,q) es conocida.
- Este hecho se puede cuantificar como la entropía condicional de p con respecto a q y que denotamos anteriormente como

$$\mathrm{H}(\Gamma;\mathrm{p})=H(p|q) \text{ cuando } \mathrm{q}=\mathrm{p}\Gamma.$$

Definición 8 (Secreto Perfecto)

Supongamos que un criptosistema es definido por un conjunto de claves K y las funciones de cifrado $E_k : M \to C$ *para* $k \in K$. *Supongamos que tenemos una distribución de probabilidad r en K. Entonces el sistema posee un secreto perfecto si*

$$H(p) = H(\Gamma; p), \text{ para todas las distribuciones p sobre } M$$

donde Γ *es la matriz del canal para la distribución r.*

Teorema 1

Un criptosistema tiene un secreto perfecto sii, para todas las distribuciones de probabilidad p en el conjunto de texto en claro M, la matriz del canal Γ para la distribución r sobre le conjunto de las claves K satisface que

$$\Gamma_{mc} = q_c \text{ para todo } m \in M y c \in C$$

donde q es la distribución correspondiente del espacio de texto cifrado C.

- El teorema sugiere que el secreto perfecto es muy raro que ocurra.
- Las entradas Γ de la matriz del canal son fijas, mientras que las probabilidades q_c varían con la distribución de entrada p.
- Por tanto la igualdad $\Gamma_{mc} = q_c$ solamente se da en muy especiales circunstancias.

Lema 2

Si un criptosistema tiene un secreto perfecto entonces el número de claves tiene que ser al menos igual al número de mensajes: $|K| \geq |M|$.

One time pad

- El secreto perfecto es una condición muy restrictiva, y los ejemplos son difíciles de obtener. Pero hay un ejemplo muy importante conocido como *one-time pad*.
- En este sistema los conjuntos de texto en claro M y de texto cifrado son ambos conjuntos de n-tuplas en un alfabeto dado, tal como el alfabeto español $\mathbb{A}$.
- Ya que la teoría es aplicable para cualquier alfabeto, para los propósitos de exposición utilizaremos el alfabeto binario $\mathbb{F}_2$.
- Tomaremos $M = C = \mathbb{F}_2^n$, el vector de cadenas de longitud n en $\mathbb{F}_2$.
- El conjunto de todas las claves K es también $\mathbb{F}_2^n$, por lo que cumplimos la condición anterior. Para cada $k \in K$ la función de cifrado E_k es

$$m \mapsto m + k (m \in M).$$

- La correspondiente función de descifrado es la misma función, $c \mapsto c + k$, ya que

$$(\mathrm{m} + \mathrm{k}) + \mathrm{k} = \mathrm{m}.$$

- Claramente tenemos un sistema de clave simétrico.
- La idea detrás del sistema es que el texto en claro es expresado como una cadena m de n bits, y es cifrado añadiendole una cadena de clave arbitraria k de la misma longitud.
- El siguiente teorema dice que el sistema tiene un secreto perfecto si la cadena de la clave es escogida uniformemente aleatoria.
- En otra palabras, cada nuevo mensaje es cifrado utilizando una nueva clave, donde todas las claves son igualmente probables, de hay la razón de su nombre.

Teorema 2

El criptosistema descrito anteriormente, con la probabilidad de distribución sobre las claves definida por

$$r_k = 1/2^n (k \in K),$$

tiene un secreto perfecto.

Métodos Iterativos

- Durante el siglo veinte los avances en los fundamentos matemáticos de la criptografía fueron de la mano con los desarrollos tecnológicos.
- Utilizando máquinas para mezclar el texto en claro de forma eficiente, se esperaba que era posible alcanzar cualquier nivel deseado de seguridad. Desafortunadamente, esa esperanza no fue alcanzada.
- La capacidad para concebir mecanismos complejos para el cifrado y descifrado se adecua a la capacidad de los mecanismos concebidos que pueden atacar y romper los sistemas resultantes.
- La conclusión obtenida fue que los criptosistemas de clave simétrica puede solamente proporcionar seguridad de forma relativa. Los sistemas pueden solamente garantizar la seguridad contra los métodos de ataque conocidos.
- En la parte tecnológica, el desarrollo más significante ha sido la aplicación de ordenadores, ya que son ideales para los cálculos que involucran iteraciones, y describiremos una procedimiento que toma ventaja de este hecho.

- Supongase que dado un trozo de texto X, expresado como una cadena de n bits, y una clave k, como una cadena de s bits.
- Sea F una función que asigna a cada X y k otra cadena de n bits, $Y = F(k, X)$. En otras palabras tenemos una función

$$F : \mathbb{F}_2^s \times \mathbb{F}_2^n \to \mathbb{F}_2^n$$

- Por razones de seguridad, F debe tener unas características especiales, tal como no ser una función lineal.
- Cuando s=2 y n=3 podríamos utilizar la función definida por la clave $k = \alpha\beta$ y $X = x_1x_2x_3$ por la regla

$$y_1 = \alpha x_1x_2 + \beta,\ y_2 = x_2 + \beta x_3\ ,\ y_3 = (\alpha + \beta)x_1x_3.$$

- Para una clave dada, esta función no es lineal, ni es una biyección.

Definición 9 (Iteración Feistel)

La iteración Feistel asociada con F y los valores iniciales X_0, $X_1 \in \mathbb{F}_2^n$, *es definida como sigue.*

Sea $k = (k_1, k_2, \ldots, k_r)$ *sea una secuencia de claves, cada una de las cuales es un elemento de* $\mathbb{F}_2^s$, *y define* $X_{i+1} = X_{i-1} + F(k_i, X_i)$ $(i = 1, 2, \ldots, r)$,

donde $+$ *representa la operación de suma bit a bit en el espacio de vectores* $\mathbb{F}_2^n$.

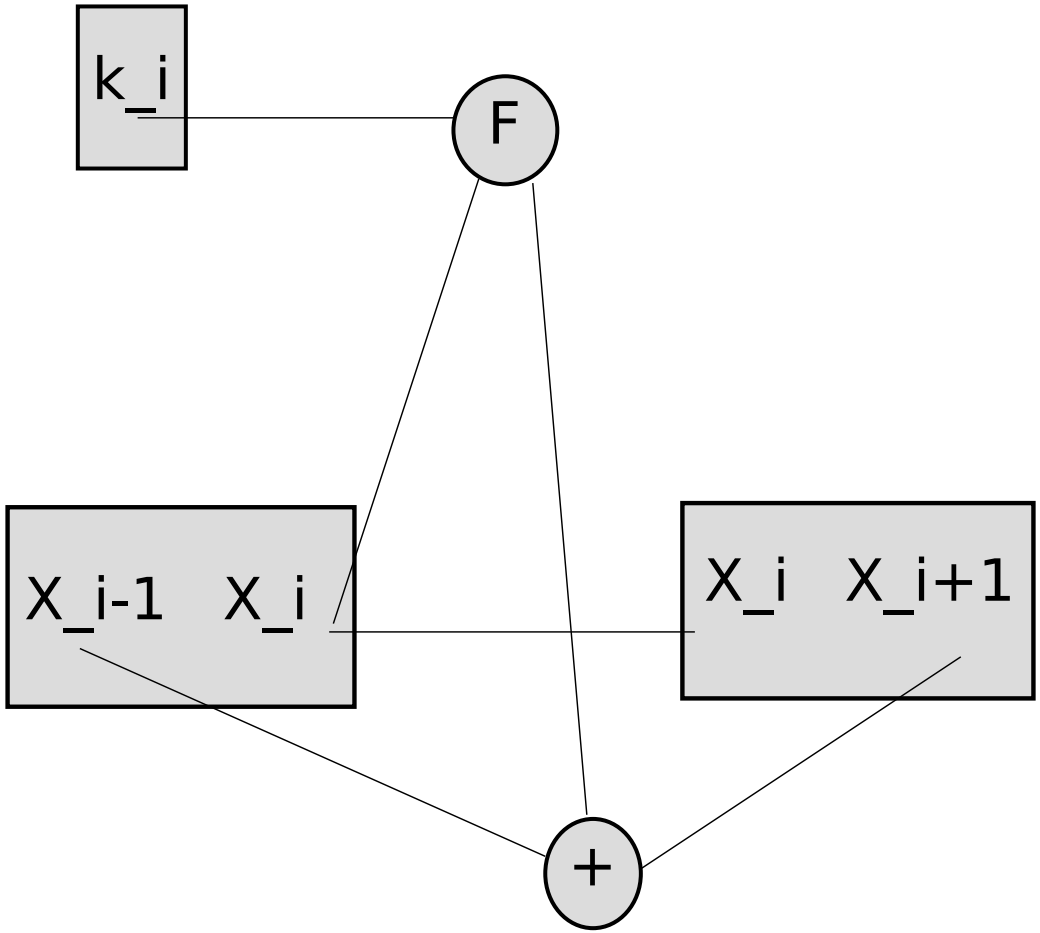

Figura 5 : La ith ronda de una iteración Feistel

- En criptografía cada paso de la iteración es denominado una ronda. En la ith ronda la clave k_i es utilizada para transformar el par $X_{i-1}X_i$ en X_iX_{i+1} (Figura 5).
- El cifrado es realizado expresando todos los mensajes como secuencias de cadenas de bits de longitud $2n$.
- Si el texto en claro contiene m $2n$-cadenas, es dividida en dos partes iguales, $m = X_0X_1$, y entonces las n-cadenas X_0 y X_1 son tomadas como los valores iniciales para la iteración de Feistel.
- El correspondiente texto cifrado es la salida después de r rondas: $c = X_rX_{r+1}$.
- Por tanto la función de cifrado para una secuencia de claves k es dada por

$$E_k(m) = c, \text{ donde } m = X_0X_1,\ c = X_rX_{r+1}.$$

Ejemplo 8

Tomaremos s=n=3 y definimos Y=F(k,X) para una clave $k = \alpha\beta\gamma$ *por la regla*

$$y_1y_2y_3 = F(\alpha\beta\gamma, x_1x_2x_3), \text{ donde}$$
$$y_1 = \alpha x_1 + x_2x_3,\ y_2 = \beta x_2 + x_1x_3,\ y_3 = \gamma x_3 + x_1x_2.$$

Si la clave es k = (100,101,001,111) y m = 101 110, calcula $c = E_k(m)$.

Solución:
Los cálculos pueden ser tabulados como sigue:

i	k_i	X_i	$F(k_i, X_i)$	$Xi+1$
0	-	101	-	110
1	100	110	101	000
2	101	000	000	110
3	001	110	001	001
4	111	001	001	111

Entonces c = 001 111.

La iteración de Feistel proporciona un mecanismo para mezclar los datos, ya que los parámetros n,s,r pueden ser tan largos como deseemos, sin embargo es un sistema de clave simétrica trivial.

Teorema 3

Supongase que el mensaje $m = X_0X_1$ *es cifrado utilizando el sistema Feistel con una función F y la secuencia de claves* $k = (k_1, k_2, \ldots, k_r)$*, por lo que* $c = E_k(m) = X_rX_{r+1}$.

Entonces el sistema Feistel con la misma función F y la secuencia de claves $k* = (k_r, k_{r-1}, \ldots, k_1)$ *cuando lo aplicamos a* $c' = X_{r+1}X_r$ *obtiene* $m' = X_1X_0$.

- Hemos notado que la función F no puede tener ciertas propiedades. Por ejemplo, si es una función lineal, el sistema Feistel será vulnerable a un ataque de texto en claro conocido.
- Como siempre, la seguridad del sistema también dependerá del tamaño de las claves usadas para evitar ataques de búsqueda exhaustiva.
- Todas las claves $k_1, k_2, \ldots k_r$ pueden ser escogidas independientemente, pero en práctica es usual extraerlas eligiendo unos conjuntos específicos de s bits de una clave maestra K.
- Por ejemplo, si necesitamos 12 claves de longitud 32, Alice y Bob pueden ponerse de acuerdo en una clave maestra de longitud 120, y tomar k_1 como los bits desde 1 a 32, k_2 como los bits 9 a 40, k_3 como los bits 17 al 48, y así continuadamente.

Estándar de Cifrado

- En los 70s la popularidad y el incremento en el uso de procedimientos criptográficos en comercio y la industria dieron lugar a necesidad de establecer un estándar de cifrado.
- Nuestra discusión anterior nos indica que cualquier criptosistema solamente debe depender para su seguridad de mantener la clave en secreto. Un sistema que pretenda ser un estándar de seguridad debe ser capaz de responder estas dos cuestiones básicas.
  - ¿Es posible el ataque por búsqueda exhaustiva?
  - ¿Existe un ataque que es mejor que la búsqueda exhaustiva?

- En 1977 se estableció el primer Estándar de Cifrado de Datos conocidos como DES.

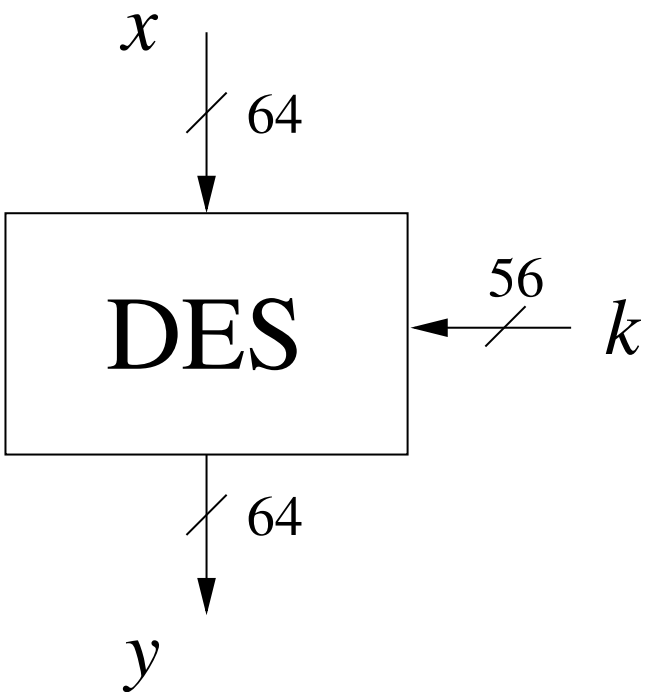

Figura 6 : Estructura General DES

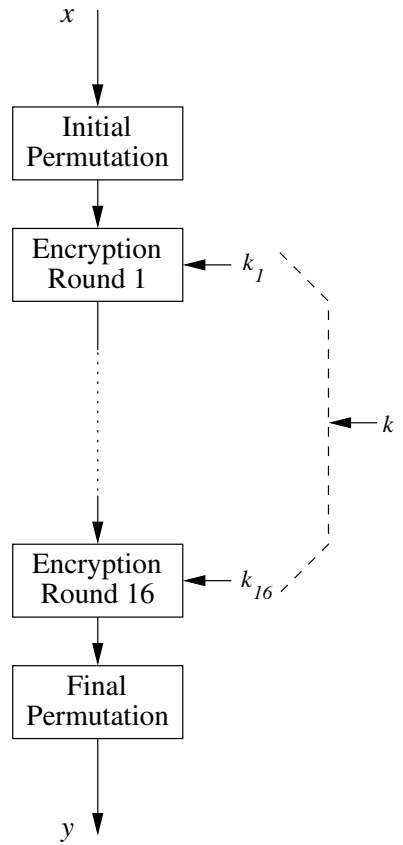

Figura 7 : Estructura General DES Ampliada

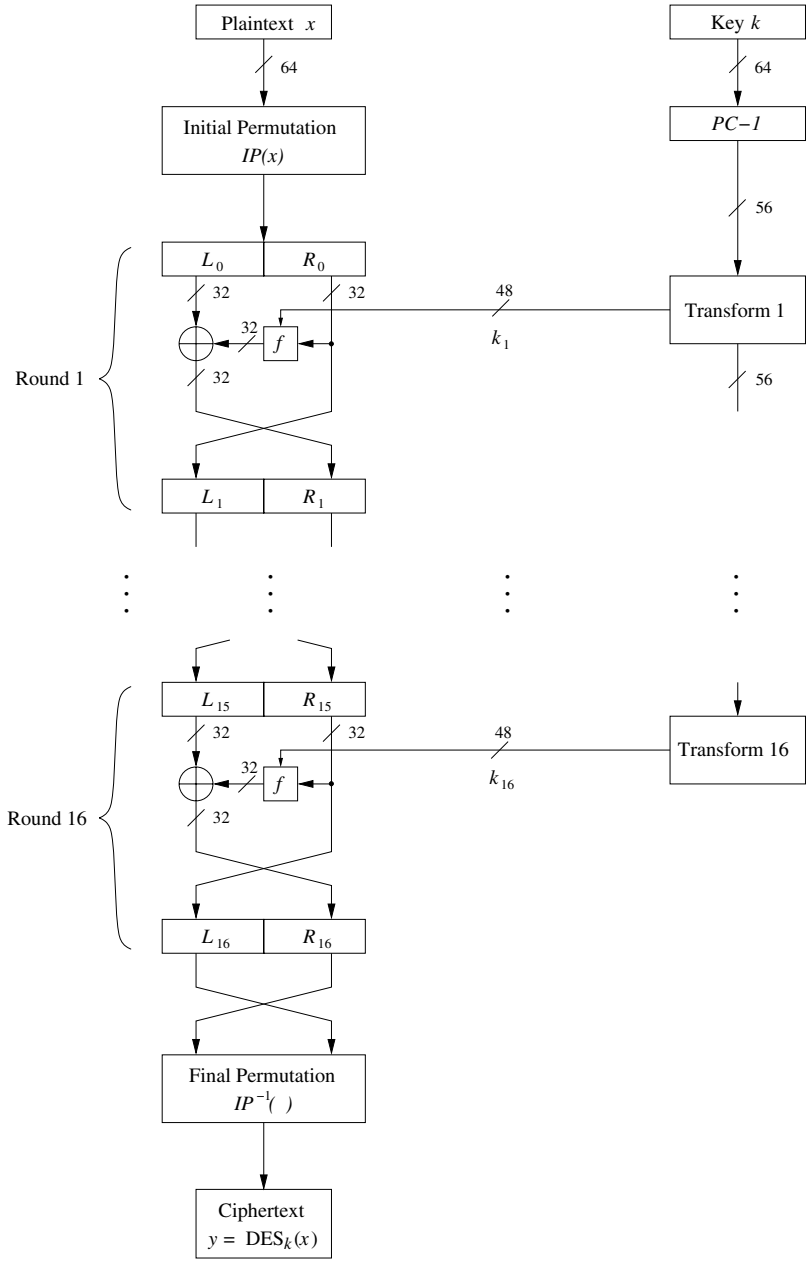

Figura 8 : Red de Feistel

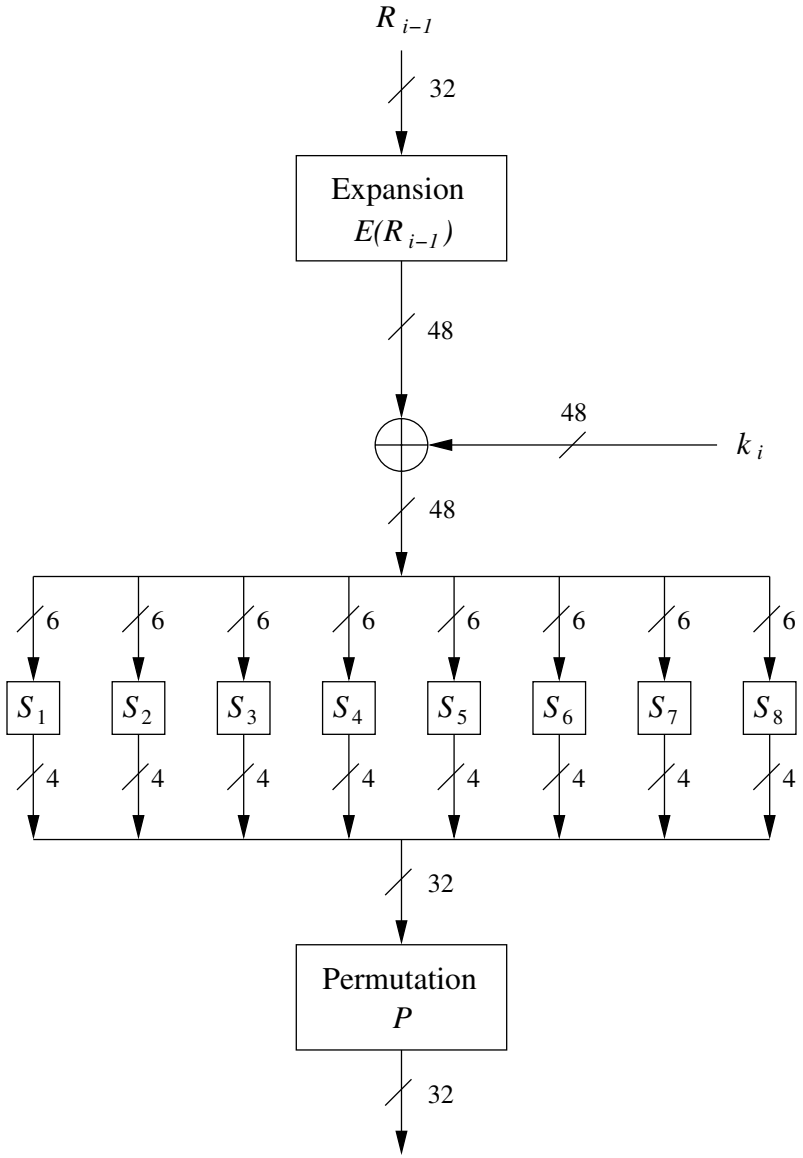

Figura 9 : Función F

S-boxes

- El componente central de la propuesta es conocido como S-box.
- En términos generales, es una matriz S donde las filas corresponden a elementos de $\mathbb{F}_2^M$, las columnas corresponden a los elementos de $\mathbb{F}_2^N$ y las entradas a $\mathbb{F}_2^R$.
- En DES hay 8 S-boxes, con los parámetros M=2, N=4 y R=4. Un ejemplo es mostrado en la Figura 11.
- Para clarificar, las filas son etiquetadas 0-3 como números binarios, $0 = 00$, $1 = 01$, $2 = 10$, $3 = 11$. Las etiquetas de las columnas y las entradas de S son denotadas por 0-15 de la misma forma: por tanto $5 = 0101$, $12 = 1100$, y así con todas.

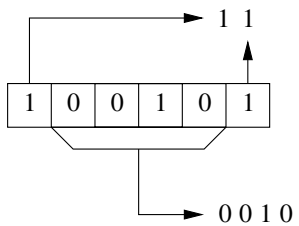

Figura 10 : Lógica funcionaniento SBox

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	12	1	10	15	9	2	6	8	0	13	3	4	14	7	5	11
1	10	15	4	2	7	12	9	5	6	1	13	14	0	11	3	8
2	9	14	15	5	2	8	12	3	7	0	4	10	1	13	11	6
3	4	3	2	12	9	5	15	10	11	14	1	7	6	0	8	13

Figura 11 : S-box Utilizada en DES

- Cada S-box en DES determina una función $f_S : \mathbb{F}_2^6 \to \mathbb{F}_2^4$ definida por la regla

$$f_S(x_1x_2x_3x_4x_5x_6) = S(x_1x_6, x_2x_3x_4x_5).$$

- Por ejemplo, para la S-box dada arriba,

$$f_S(111001) = S(11, 1100) = S(3, 12) = 6 = 0110.$$

- La función de cifrado DES opera con bloques de 64 bits, utilizando una clave con (esencialmente) 56 bits.
- Es obtenida mediante permutación de los bits y combinando las 8 funciones f_S definidas por las S-boxes, utilizando una iteración de Feistel con 16 rondas.

- Las S-boxes son construidas de forma que satisfagan un número de criterios, diseñadas para hacerlas más segura frente métodos conocidos de ataques. Dos de esos criterios son:

  C1 Cada fila de S es una permutación $\mathbb{F}_2^4$
  C2 Si x,y $\in \mathbb{F}_2^6$ cumple que la distancia Hamming d(x,y)=1, $yx' = f_S(x)$, $y' = fS_(y)$, entonces $d(x', y') \geq 2$.

- La propuesta del DES estableció mucha controversia.
- El método de construir las S-boxes no fuer revelado, un hecho que algunas personas tildan de sospechoso.
- Los parámetros escogidos son muy pequeños, principalmente porque la clave maestra solo tiene 56 bits efectivos, esta situación no libera de un ataque por búsqueda exhaustiva.

SubClaves DES

- En las rondas 1,2,9,16 las dos partes son rotadas un bit hacia la izquierda.
- En las restantes rondas 3,4,5,6,7,8,10,11,12,13,14,15 las partes son rotadas dos bits.

El número total de rotaciones son $4 \times 1 + 12 \times 2 = 28$.
Esta hecho da lugar a que $C_0 = C_{16}$ y $D_0 = D_{16}$.

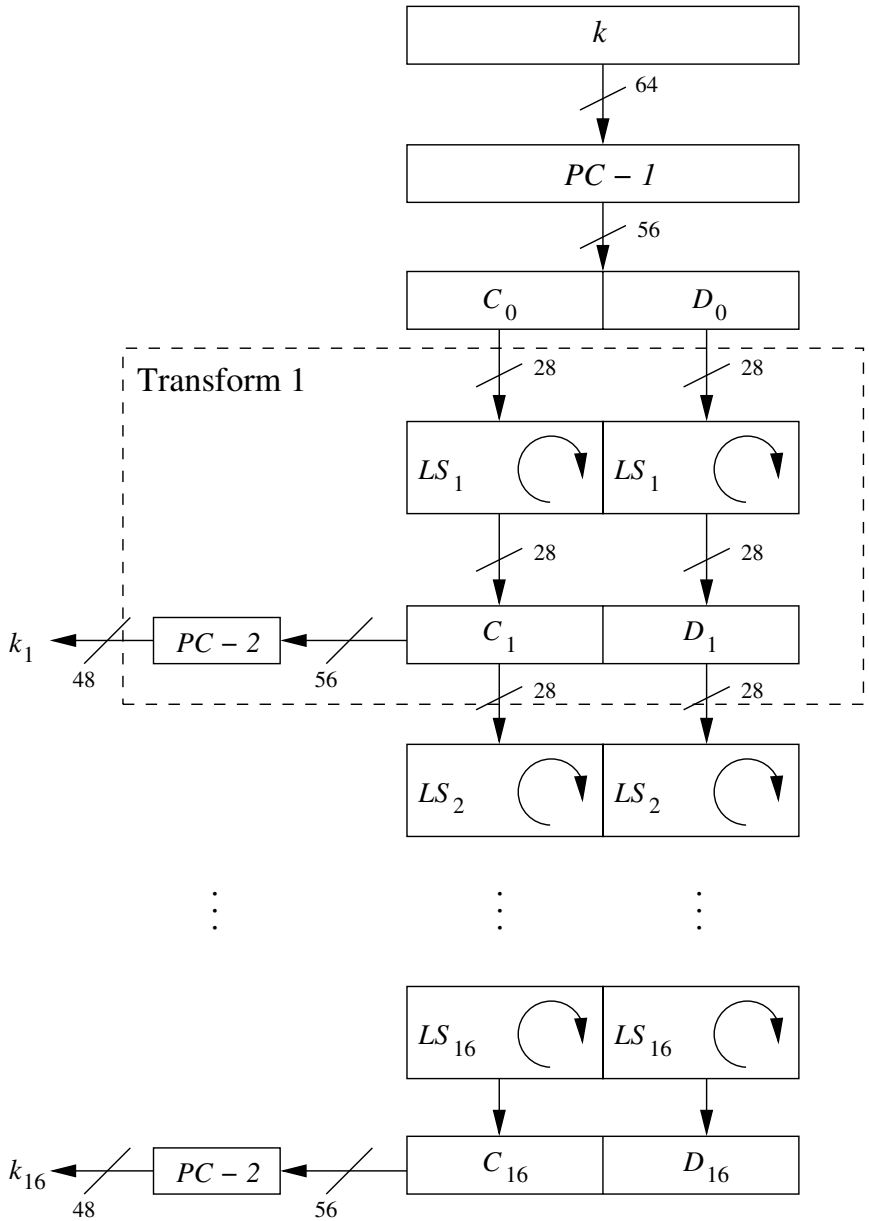

Figura 12 : Subclaves Cifrado

$PC-1$							
57	49	41	33	25	17	9	1
58	50	42	34	26	18	10	2
59	51	43	35	27	19	11	3
60	52	44	36	63	55	47	39
31	23	15	7	62	54	46	38
30	22	14	6	61	53	45	37
29	21	13	5	28	20	12	4

Figura 13 : PC1

$PC-2$							
14	17	11	24	1	5	3	2
15	6	21	10	23	19	12	4
26	8	16	7	27	20	13	2
41	52	31	37	47	55	30	4
51	45	33	48	44	49	39	5
34	53	46	42	50	36	29	3

Figura 14 : PC2

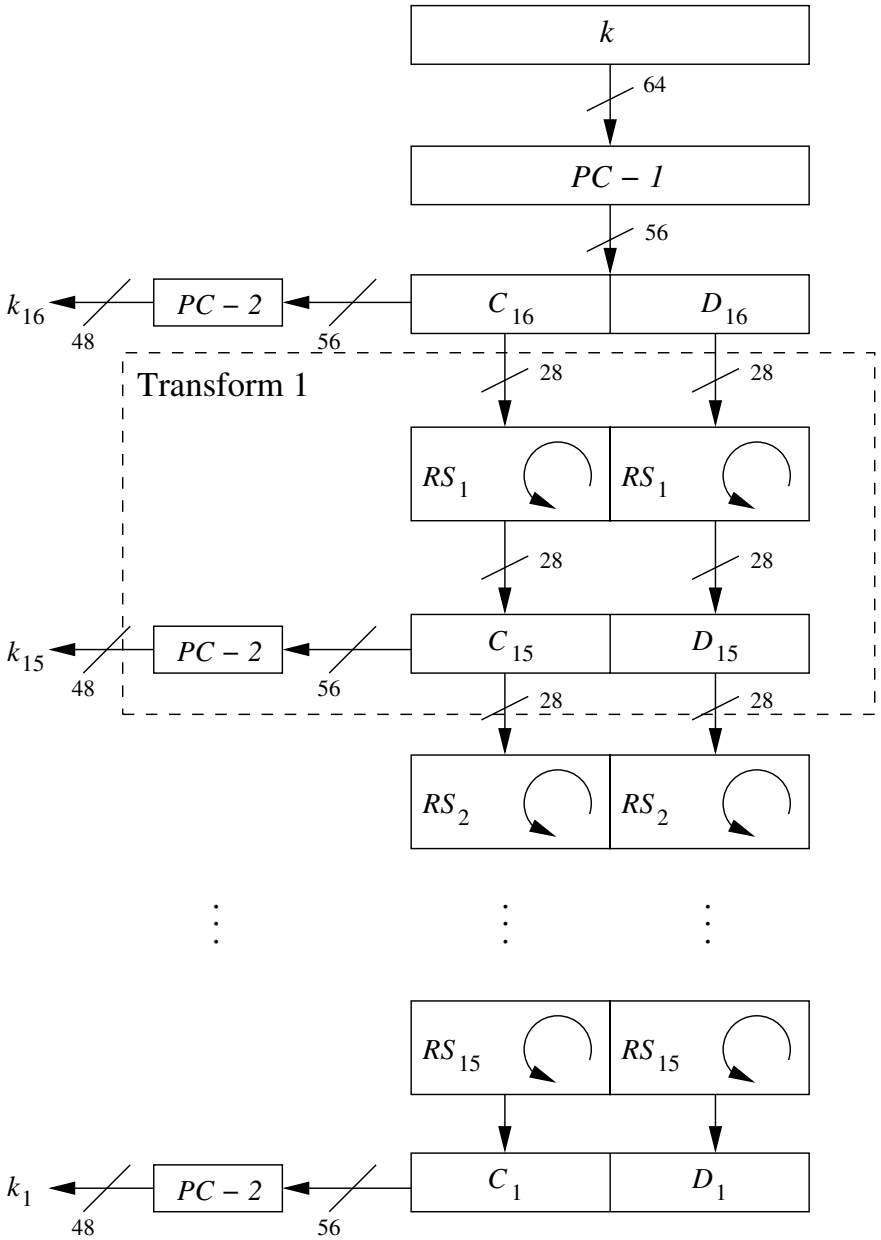

Figura 15 : Subclaves Descifrado

- Debida a esta situación, en los 90s, estaba claro que DES no podía contestar estas cuestiones satisfactoriamente, por tanto fue reemplazado por el estándar de cifrado avanzado o AES.
- En este sistema la clave maestra tiene 128 bits, y actualmente nos es posible llevar un ataque por búsqueda exhaustiva.
- El sistema no está basado directamente en iteraciones de Feistel, pero principios similares son aplicados.
- Hay solamente una S-box, con M=N=4 y R=8 y es definido mediante una fórmula.

El problema de distribución de claves

- En un criptosistema de clave simétrica, Alice y Bob utilizan claves similares.
- Metafóricamente hablando, Alice pone el mensaje en una caja la cual ella cierra utilizando la clave k.
- Cuando Bob recibe la caja la abre utilizando una clave l que está muy relacionada con k, podemos establecer esta relación como $l = k'$.
- El principal problema es que Alice y Bob deben comunicarse para ponerse de acuerdo en la clave, y esta comunicación no puede ser por tanto protegida por el proceso de cifrado.
- Esta situación se conoce como el problema de distribución de claves.

- Hay varias formas en las cuales el problema de distribución de claves puede ser solucionado.
- Si tenemos recursos disponibles, podríamos utilizar el one-time pad, de esta manera antes de transmitir cada mensaje Alice envía una nueva clave a Bob de forma segura.
- Claramente, este caso no es práctico para la mayoría de os propósitos. Una medida más práctica es considerar alternativas a los procedimientos de clave simétrica.
- ¿Es posible designar un procedimiento en el cual la clave de Alice y Bob sean independientes, de forma que no necesiten conocer las claves de uno del otro?

Un posible procedimiento sería el descrito a continuación:

1. Alice envía a Bob un mensaje cifrado utilizando su clave a.
2. Bob cifra el mensaje utilizando su clave b y lo devuelve a Alice.
3. Alice descifra utilizando a'y lo envía a Bob.
4. Bob descifra utilizando b'.

El procedimiento funciona si las operaciones de Alice conmutan con las de Bob.

El Criptosistema RSA

- En la sección anterior hemos tratado el problema intrínseco que presentan los criptosistemas de clave simétrica debido al problema de la distribución de las claves.
- En 1970s, el desarrollo de la criptografía de clave pública vino a solventar la cita situación.
- La idea fundamental es que un usuario (Bob) tiene dos claves, una pública y otra privada.
- La clave pública es utilizada por Alice y otros para cifrar mensajes que ellos quieren enviar a Bob, y la clave privada es utilizada por Bob para descifrar estos mensajes.
- La seguridad del sistema recae en el aserto que la clave privada de Bob no puede ser calcula, incluso si todo el mundo conoce su clave pública.

- Un método práctico de implementar esta idea fue descubierta por Rivest, Shamir, y Adleman en 1977 y es conocida como criptosistema RSA.
- En el sistema RSA los textos en claro son secuencias de enteros $mod\, n$, donde es razonable determinar n como un número de al menos 300 dígitos decimales.

Definición 10 (Funcion ϕ)

Para cualquier entero positivo n, el número de enteros x en el rango $1 \leq x \leq n$ tal que mcd(x,n)=1 es denotado como $\phi(n)$. $\phi(n)$ representan también el número de elementos invertibles de $\mathbb{Z}_n$.

- Por ejemplo tomando n=14 los valores de x tal que mcd(x,14) $= 1$ son 1,3,5,9,11,13 por lo que $\phi(14) = 6$.
- En este caso las inversas son fácilmente calculables, por ejemplo $3^{-1} = 5$ ya que $3 \times 5 = 1$ mod 14.
- Si n es un número primo, $n = p$, entonces los p-1 números $x = 1, 2, \ldots, p-1$ satisfacen mcd(x,p) $= 1$ y por tanto $\phi(p) = p - 1$.

Lema 3

Si $n = pq$, donde p y q son primos, entonces $\phi(n) = (p-1)(q-1)$.

En el sistema RSA, los usuarios deben de realizar los siguientes pasos previos a la realización de las funciones de cifrado y descifrado:

- Escoge dos números primos p,q y calcula

$$n = p\,q,\ \phi = (p-1)(q-1)$$

- Escoge e tal que $mcd(e, \phi) = 1$, y calcula

$$d = e^{-1} (mod\, \phi)$$

Las funciones de cifrado y descifrado son definidas como sigue:

$$E_{n,e}(m) = m^e \ (m \in \mathbb{Z}_n),$$
$$D_{n,d}(c) = c^d \ (c \in \mathbb{Z}_n)$$

El sistema funciona de la siguiente forma:

- Bob utiliza las reglas anteriores para construir, los números n, ϕ, e, y d.
- Una vez calculados los elementos, pone a disposición de todos la *clave pública (n,e)*, pero mantiene la *clave privada d* en secreto.
- Cuando Alice quiere enviar a Bob un mensaje, lo expresa inicialmente en una secuencia de enteros m mod n, calcula $c = E_{n,e}(m)$ y envía c.
- Bob utiliza su clave privada para calcular $m = D_{n,d}(c)$.

Ejemplo 9

Supongamos Bob ha elegido $p = 47$, $q = 59$.
(i) Encontrar n y ϕ, y muestra que e=157 es una valor válido para la clave pública.
(ii) Verifica que su clave privada es d=17.
(iii) Explicar los pasos que debe realizar Alice para enviar un mensaje cifrado a Bob representado por una secuencia de enteros, por ejemplo m=5. ¿Como Bob lo podría descifrar?

Solución:

(i) Tenemos

$$n = 47 \times 59 = 2773,\ \phi = 46 \times 58 = 2668.$$

La elección e $= 157$ es válida ya que mcd(2668, 157) $= 1$. Esta condición se cumple ya que 157 es primo y no divide 2668.
(ii) Podemos probar la condición calculando:

$$d \times e = 17 \times 157 = 2669 = 1 \text{ (mod } 2668).$$

continua ...

(iii) Si Alice quiere enviar a Bob un mensaje, buscará su clave pública (n,e)=(2773, 157), y realizará un proceso de conversión del mensaje en una secuencia de enteros mod n. Después cifrará el mensaje utilizando la función de cifrado $E_{n,e(m)} = m^e$, y envía el mensaje cifrado a Bob. En este caso

$$c = 5^{157} = 1044\,(mod\,2773).$$

Cuando Bob recibe el texto cifrado c = 1044, aplicará su función de descifrado $D_{n,d(c)} = c^d$, obteniendo

$$m' = 1044^{17} = 5 \text{ (mod 2773)}.$$

Podemos probar que $m' = m$, por lo que el mensaje original es obtenido.

Viabilidad de RSA

Daremos una breve explicación de como realizar los cálculos en RSA fácilmente. Básicamente depende de la ejecución de dos algoritmos, los cuales son usados para reducir problemas complicados a una aritmética básica.

- El algoritmo de Euclides es esencialmente un método para calcular el mcd(a, b) de dos enteros a y b (asumimos que $a < b$). Depende básicamente del hecho que si

$$b = q\,a + r \text{ entonces } mcd(a, b) = mcd(r, a).$$

- Podemos remplazar (a, b) por $(a', b') = (r, a)$, y repetir el proceso.

Ejemplo 10

Encuentra el mcd(654,2406).

Ejemplo 10

Encuentra el mcd(654,2406).

Solución:

El proceso se realiza como sigue:

$$\begin{aligned} 2406 &= 3 \times 654 + 444 \\ 654 &= 1 \times 444 + 210 \\ 444 &= 2 \times 210 + 24 \\ 210 &= 8 \times 24 + 18 \\ 24 &= 1 \times 18 + 6 \\ 18 &= 3 \times 6. \end{aligned}$$

Por tanto 6 es el máximo común divisor que divide 18, 24, 210, 444, 654, 2406, y en general mcd(654, 2406) $= 6$.

- Si mcd(a, b) $= 1$, entonces a tiene una inversa multiplicativa mod b, y el algoritmo de Euclides puede ser utilizado para calcularlo.
- El algoritmo es aplicado de forma inversa de la forma $\lambda\, a + \mu\, b$, donde λ y μ son enteros.
- La ecuación $1 = \lambda\, a + \mu\, b$ puede ser reescrita de la forma $\lambda\, a = 1 (mod\, b)$. Por tanto λ es la inversa de a (mod b).

Ejemplo 11

Encontrar la inversa de 24 mod 31.

Ejemplo 11

Encontrar la inversa de 24 mod 31.

Solución:

Primero verificamos que mcd(24, 31) $= 1$:

$$31 = 1 \times 24 + 7$$
$$24 = 3 \times 7 + 3$$
$$7 = 2 \times 3 + 1$$
$$3 = 3 \times 1.$$

Ahora, comenzando con la última ecuación y volviendo hacia atrás tenemos,
$1 = 7 - 2 \times 3$
$= 7 - 2 \times (24 - (3 \times 7)) = 7 \times 7 - 2 \times 24$
$= 7 \times (31 - 1 \times 24) - 2 \times 24 = -9 \times 24 + 7 \times 31.$
Por tanto $(-9) \times 24 = 1$ (mod 31), por lo que $24^{-1} = -9 = 22$.

Ejemplo 12

En el ejemplo anterior Bob establece una clave RSA eligiendo $p = 47$, $q = 59$, por lo que $n = 2773$, $\phi = 2668$. Utiliza el algoritmo de euclides para demostrar que $e = 157$ es valor válido para su clave pública, y explica como puede calcular su clave privada $d = e^{-1}(mod\,\phi)$.

Ejemplo 12

En el ejemplo anterior Bob establece una clave RSA eligiendo $p = 47$, $q = 59$, por lo que $n = 2773$, $\phi = 2668$. Utiliza el algoritmo de euclides para demostrar que $e = 157$ es valor válido para su clave pública, y explica como puede calcular su clave privada $d = e^{-1}(mod\,\phi)$.

Solución:

El algoritmo de Euclides muestra que $mcd(e, \phi) = 1$:

$$\begin{aligned} 2668 &= 16 \times 157 + 156 \\ 157 &= 1 \times 156 + 1 \\ 156 &= 156 \times 1. \end{aligned}$$

$d = e^{-1}$ es calculada:

$$\begin{aligned} 1 &= 157 - 1 \times 156 \\ &= 157 - 1 \times (2668 - (16 \times 157) \\ &= (-1) \times 2668 + 17 \times 157. \end{aligned}$$

Por tanto $17 \times 157 = 1(mod\,2668)$, so por lo que 17 es la clave privada de Bob.

- El otro algoritmo utilizado tiene que ver con las funciones exponenciales $a^b \, (mod \, c)$ utilizadas en las funciones de cifrado $(c = m^e)$ y descifrado $(m = c^d)$.
- Básicamente para calcular la potencia b^k debemos de utilizar k-1 multiplicaciones, ya que $b^k = b \times b \times \ldots \times b$.
- Sin embargo el algoritmo de exponenciación modular o de cuadrados repetido nos proporciona una mejor solución para realizar los cálculos.

Ejemplo 13

Supongamos que dado el número b. ¿Cuantas multiplicaciones necesitamos para calcular b^{16} y cuantas para b^{23}?

Ejemplo 13

Supongamos que dado el número b. ¿Cuantas multiplicaciones necesitamos para calcular b^{16} y cuantas para b^{23}?

Solución:

El cálculo de b^16 requiere solamente cuatro multiplicaciones:
$b^2 = b \times b$, $b^4 = b^2 \times b^2$, $b^8 = b^4 \times b^4$, $b^{16} = b^8 \times b^8$.
Sabiendo que $23 = 16 + 4 + 2 + 1$, solamente necesitamos tres multiplicaciones más para b^{23}:
$b^{23} = b^{16} \times b^4 \times b^2 \times b$.
Por tanto b^{23} puede ser realizadas con 4+3=7 multiplicaciones.

Fiabilidad de RSA

Para explica como funciona RSA, nos basamos es explicar porqué la función $D_{n,d}$ es la inversa de $E_{n,e}$.
Acorde con la definición, la condición

$$D_{n,d}(E_{n,e}(m)) = m$$

reduce a $(m^e)^d =$ m (mod n), por lo que debemos probar que esta situación se da cuando d es la inversa de e mod $\phi(n)$.

Lema 4

Si mcd(x,n) = 1, por lo que x tiene una inversa en $\mathbb{Z}_n$, entonces $x^{\phi(n)} = 1$ en $\mathbb{Z}_n$.

Teorema 4

Si las funciones de cifrado y descifrado RSA son $E_{n,e}$ y $D_{n,d}$, y mcd(m, n) = 1, entonces

$$D_{n,d}(E_{n,e}(m)) = m.$$

O lo que es lo mismo, $m^e = m(mod\, n)$.

Ejercicio 1

Realiza una lista de aquellos elementos de $\mathbb{Z}_{36}$ que tiene inversa multiplicativas, y encuentra la inversa de 5 mod 36.

Ejercicio 1

Realiza una lista de aquellos elementos de $\mathbb{Z}_{36}$ que tiene inversa multiplicativas, y encuentra la inversa de 5 mod 36.

Solución:

El número de elementos es 12, y $5^{-1} = 29$.

Ejercicio 2

Evalúa $\phi(257)$ *y* $\phi(253)$

Ejercicio 2

Evalúa $\phi(257)$ *y* $\phi(253)$

Solución:

$\phi(257) = 256$ y $\phi(253) = 220$

Ejercicio 3

Jorge y Antonio utilizan el criptosistema RSA. Jorge publica su clave pública $n = 77$, $e = 43$.¿Cuáles son los primos p y q, y cuál es su clave privada?. Antonio envía un mensaje $m \in \mathbb{Z}_n$, y Pepe intercepta el mensaje cifrado c=5. ¿Cuál era m?

Ejercicio 3

Jorge y Antonio utilizan el criptosistema RSA. Jorge publica su clave pública n = 77, e = 43.¿Cuáles son los primos p y q, y cuál es su clave privada?. Antonio envía un mensaje $m \in \mathbb{Z}_n$, y Pepe intercepta el mensaje cifrado c=5. ¿Cuál era m?

Solución:

En este caso n=77 por lo que es muy fácil factorizar en $77 = 11 \times 7$, eso nos da que por ejemplo $p = 7$ y $q = 11$.
De esta información sabemos el valor de ϕ ya que
$\phi = (p-1) \times (q-1) = 6 \times 10 = 60$, también sabemos $d = e^{-1} mod\, \phi$, por lo que $d = 43^{-1} mod\, 60 = 7$.
Pepe sabe que $m = c^d \, mod\, n$ por lo que $m = 5^7 \, mod\, 77 = 47$.

Ejercicio 4

Un usuario de RSA ha anunciado que la clave pública n = 2903239, e = 5. María ha conseguido averiguar que $n = 1237 \times 2347$. *Verifica que la clave pública es válida y explica por qué la clave privada es d = 2319725.*

Ejercicio 4

Un usuario de RSA ha anunciado que la clave pública n = 2903239, e = 5. María ha conseguido averiguar que $n = 1237 \times 2347$. Verifica que la clave pública es válida y explica por qué la clave privada es d = 2319725.

Solución:

Sabemos que $\phi = 1236 \times 2346 = 2899656$.
Calculamos mcd (2899656,5) $= 1$ que verifica que tiene una inversa.
$d \times e = 2319725 \times 5 = 11598625$ mod $2899656 = 1$.

Ejercicio 5

Utiliza el algoritmo de Euclides para mostrar que mcd(15,68)=1. Además encuentra la inversa de 15 mod 68.

Ejercicio 5

Utiliza el algoritmo de Euclides para mostrar que mcd(15,68)=1. Además encuentra la inversa de 15 mod 68.

Solución:

$15^{-1} \bmod 68 = 59$.

José A. Montenegro Montes
Dpto. Lenguajes y Ciencias de la Computación
ETSI Informática. Universidad de Málaga
monte@lcc.uma.es

UNIVERSIDAD
DE MÁLAGA

E.T.S. INGENIERÍA
INFORMÁTICA

LENGUAJES Y
CIENCIAS DE LA
COMPUTACIÓN
UNIVERSIDAD DE MÁLAGA