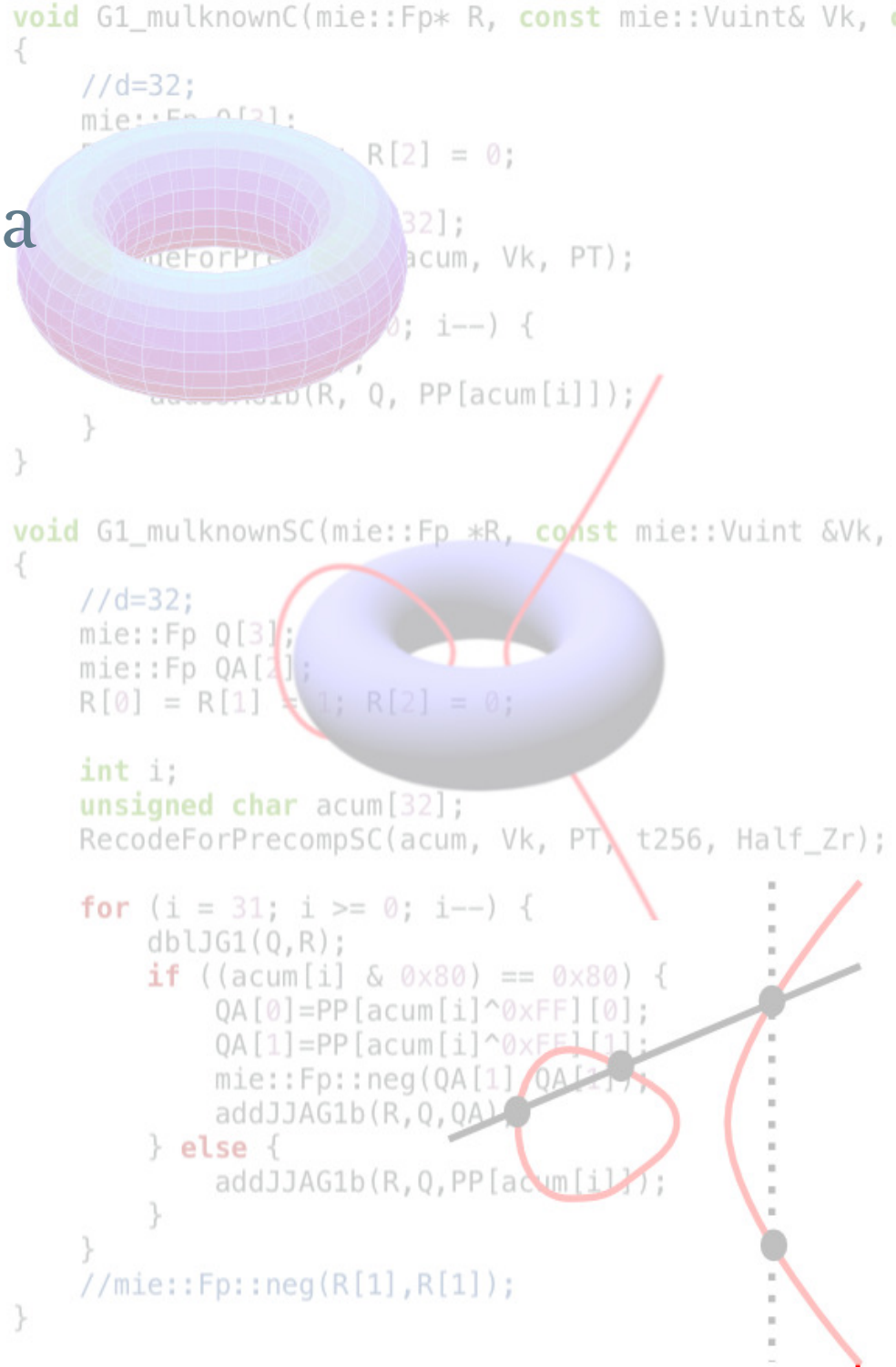
Matemáticas aplicadas a la criptografía

Unidad VI y VII - Cifradores de BLoque y de Flujo

Dr. Luis J. Dominguez Perez

Universidad Don Bosco

Abril 27--28, 2013



Contenido de la sección 1

Cifrador por sustitución

También conocido como cifrador por reemplazo, es uno de los métodos más simples para cifrar un texto.

La idea es sustituir cada letra del alfabeta por otra (o la misma), de tal manera que el texto no pueda entenderse a simple vista:

Ejemplo:

$$A \to L$$
$$B \to C$$
$$C \to J$$
$$\vdots$$

BABA = CLCL

Ataques al cifrador por sustitución

- Fuerza bruta, búsqueda exhaustiva.

El atacante tiene el texto cifrado gracias a que escuchó la conversación; además tiene una parte del texto original, por ejemplo: la cabecera del mensaje (i.e. *%PDF-1.4, PK, GIF87a, 0xFFD8*)

Ahora solo tiene que probar atacar el inicio del texto con todas claves posibles hasta que coincida.

Ataque por fuerza bruta

Formalmente,

Búsqueda exhaustiva básica de clave o ataque de fuerza bruta

Dada una pareja (x, y), el texto en claro y el texto cifrado, y sea $K = \{k_0, \ldots, k_{n-1}\}$ sea el espacio de todas las posibles claves. Un ataque de fuerza bruta verifica a todo $k_i \in K$ si:

$$d_{k_i}(y) \stackrel{?}{=} x,$$

Si la relación lógica se mantiene, entonces se ha encontrado la clave y se detiene el proceso, de otro modo, se continua.

$d(.)$ es la función de descifrado. En la práctica depende del protocolo.

Ataques por fuerza bruta

En principio, todos los cifradores *simétricos* son susceptibles a ataques por fuerza bruta. Que sea factible o no, depende del espacio de la clave (el número de posibles claves).

Por ejemplo, el NIP de las tarjetas es de 4 dígitos, existen 10^4 posibles NIPs. En este caso, robar dinero de un cajero automático tardaría nada si no fuera porque los bancos bloquean las tarjetas ante ataques.

Ataques por fuerza bruta

En cambio, si al realizar un ataque utilizando alguna computadora moderna toma mucho tiempo (i.e. décadas), se dice que el cifrador es *computacionalmente seguro* ante ataques de fuerza bruta.

En el caso del cifrador por sustitución, la letra *A* se sustituyó por la letra *L*, pero teníamos 26 opciones. La letra *B* se sustituyó por la letra *C*, de las 25 opciones restantes. Y así sucesivamente.

El número de posibles sustituciones en un ataque por fuerza bruta es:

$$26 \cdot 25 \cdots 1 = 26! \approx 2^{88}.$$

Ataques por fuerza bruta

En cambio, si al realizar un ataque utilizando alguna computadora moderna toma mucho tiempo (i.e. décadas), se dice que el cifrador es *computacionalmente seguro* ante ataques de fuerza bruta.

En el caso del cifrador por sustitución, la letra *A* se sustituyó por la letra *L*, pero teníamos 26 opciones. La letra *B* se sustituyó por la letra *C*, de las 25 opciones restantes. Y así sucesivamente.

El número de posibles sustituciones en un ataque por fuerza bruta es:

$$26 \cdot 25 \cdots 1 = 26! \approx 2^{88}.$$

¿Cuánto es 2^{88}?

- Un procesador Intel Core i7 @3.4 GHz realiza alrededor de 2^{31} ciclos de reloj por segundo, aunque tiene 4 cores...

Ciclos de reloj

- Para realizar 2^{32} ciclos de reloj, se requieren el doble de cores que el nivel anterior (2^{31}), esto es, el procesador realiza 2^{33} ciclos de reloj en total.
- tenemos $88 - 32 = 55$, hay que duplicar la cantidad de cores 55 veces.
- Es un crecimiento geométrico, tal como la leyenda del Ambalappuzha Paal Payasam (granos de arroz y el ajedrez)

Ciclos de reloj

- Para realizar 2^{32} ciclos de reloj, se requieren el doble de cores que el nivel anterior (2^{31}), esto es, el procesador realiza 2^{33} ciclos de reloj en total.
- tenemos $88 - 32 = 55$, hay que duplicar la cantidad de cores 55 veces.
- Es un crecimiento geométrico, tal como la leyenda del Ambalappuzha Paal Payasam (granos de arroz y el ajedrez)

- Sin embargo, eso es sólo 1 segundo... además de que en algunos casos hay que analizar la información.

Para fines de benchmarking, normalmente no se utiliza el TurboBoost, además que en un ataque aumentaría la radiación térmica.

Entonces se dice que este esquema es seguro ante ataques de fuerza bruta... (a menos que el *bruto* haya sido el que seleccionó la clave)

- Sin embargo, eso es sólo 1 segundo... además de que en algunos casos hay que analizar la información.

Para fines de benchmarking, normalmente no se utiliza el TurboBoost, además que en un ataque aumentaría la radiación térmica.

Entonces se dice que este esquema es seguro ante ataques de fuerza bruta... (a menos que el *bruto* haya sido el que seleccionó la clave)

- Sin embargo, eso es sólo 1 segundo... además de que en algunos casos hay que analizar la información.

Para fines de benchmarking, normalmente no se utiliza el TurboBoost, además que en un ataque aumentaría la radiación térmica.

Entonces se dice que este esquema es seguro ante ataques de fuerza bruta... (a menos que el *bruto* haya sido el que seleccionó la clave)

Un ataque *diferente*

En el ataque por fuerza bruta, tomamos al cifrador como una caja negra, sin analizarla internamente.

El cifrador por sustitución puede romperse mediante un ataque analítico.

La principal debilidad del cifrador, es que cada símbolo del texto en claro tiene una única representación en el texto cifrado. Esto es, que las propiedades estadísticas del texto en claro se preservan en el texto cifrado.

Letras en los idiomas

La letra que más se repite en el idioma inglés es la letra ``e'' (alrededor del 13% de los textos), después la ``t'' con un 9%, y la ``a'' con un 8%.

En español, la frecuencia es similar (la ``e'' también es la más utilizada). Se puede construir una tabla para el idioma español tomando cualquier libro y contando la ocurrencia de cada una de las letras.

Pero aquí están ordenadas de mayor a menor frecuencia: E A O S R N I D L C T U M P B G V Y Q H F Z J Ñ X W K

Letras en los idiomas

La letra que más se repite en el idioma inglés es la letra ``e'' (alrededor del 13% de los textos), después la ``t'' con un 9%, y la ``a'' con un 8%.

En español, la frecuencia es similar (la ``e'' también es la más utilizada). Se puede construir una tabla para el idioma español tomando cualquier libro y contando la ocurrencia de cada una de las letras.

Pero aquí están ordenadas de mayor a menor frecuencia: E A O S R N I D L C T U M P B G V Y Q H F Z J Ñ X W K

Notas sobre el conteo de letras

- En un diccionario, la letra que más se repite tiende a ser la ``a''
- En un libro, como en el Quijote, se mantiene el orden antes mencionado.
- Aunque hay excepciones.
- además, existen muchas frases cortas en el español que van con la ``e'': qué, le, sé, etc.

Finalmente, con la estadística es muy fácil descifrar un texto por sustitución.

Cifrado de Cæsar

El cifrador de César es un tipo especial del cifrador por sustitución en el cual los valores del alfabeto se rotaban una cantidad de letras en particular.

Por ejemplo, si la clave era 13, entonces la tabla de sustitución es:

$$A \to N$$
$$B \to \tilde{N}$$
$$C \to O$$
$$\vdots$$

Para el idioma español.

¿Cuál es el espacio de claves?

Cifrado de Cæsar

El cifrador de César es un tipo especial del cifrador por sustitución en el cual los valores del alfabeto se rotaban una cantidad de letras en particular.

Por ejemplo, si la clave era 13, entonces la tabla de sustitución es:

$$A \to N$$
$$B \to \tilde{N}$$
$$C \to O$$
$$\vdots$$

Para el idioma español.

¿Cuál es el espacio de claves?

Máquina enigma

La máquina enigma era una especie de máquina de escribir con un determinado número de rotores en serie que giraban de diferente forma a cada pulsación de una tecla, de tal manera que la salida de un rotor era la entrada para el siguiente, y así sucesivamente.

Estos rotores podían cambiar su posición inicial (que era la clave), de tal manera que cada vez que se enviaba un mensaje, se utilizaba una configuración diferente.

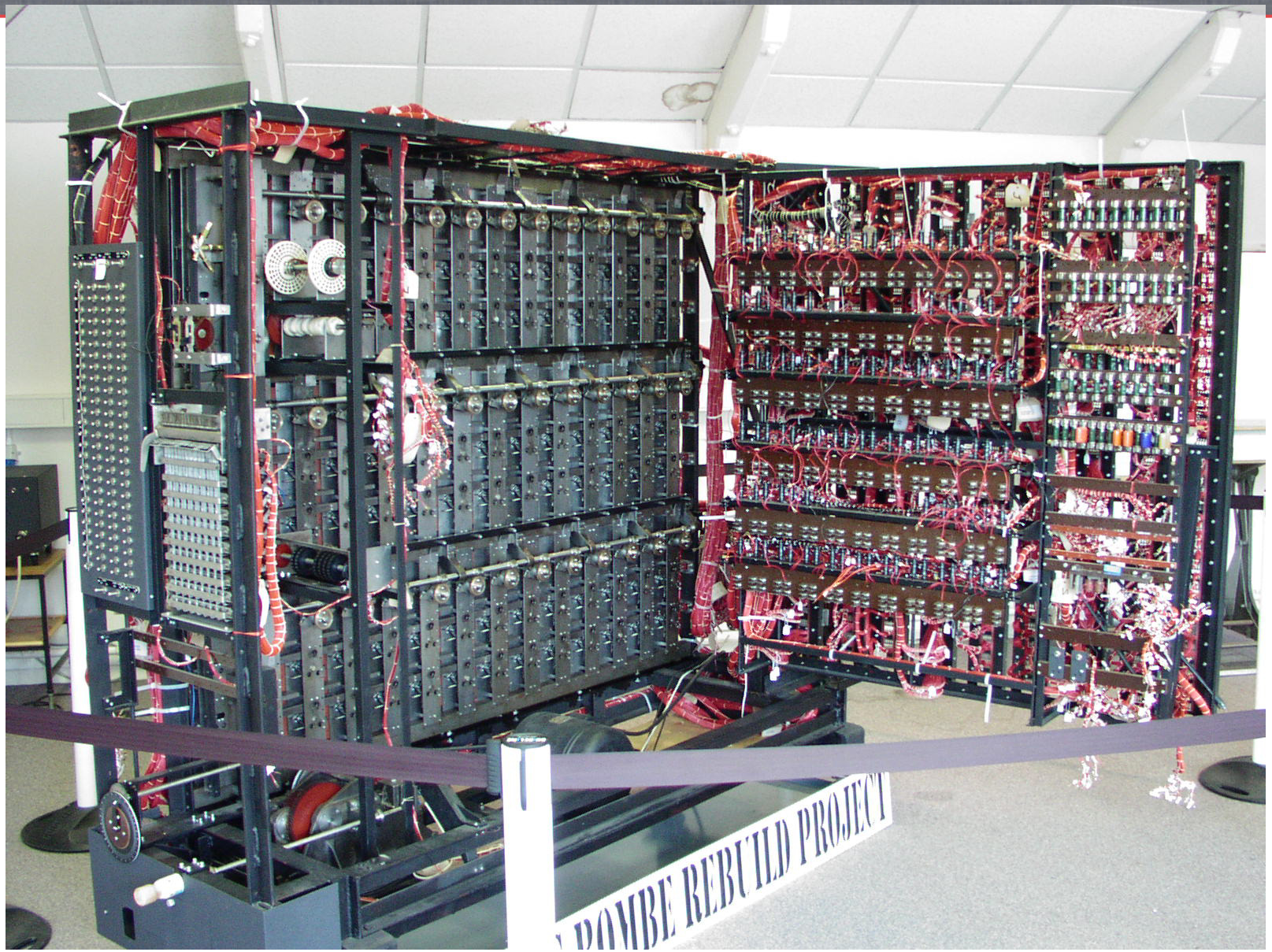
Segunda Guerra Mundial

Durante el Franquismo, los nazis proveyeron a Francisco Franco de unas máquinas limitadas para comunicarse, y probarlas.

Ya probadas, se utilizaron extensamente durante la Segunda Guerra mundial para enviar instrucciones a las líneas de batalla.

Todos los ataques alemanes eran sorpresa y 100% efectivos. Fue hasta que los británicos con su laboratorio en Bletchley Park (liderado por Alan Turín), y las máquinas polacas *bombe*, que pudieron descifrar los mensajes, y salvar millones de vidas.

Bombes polacas



Simulador máquina enigma

En línea

Contenido de la sección 2

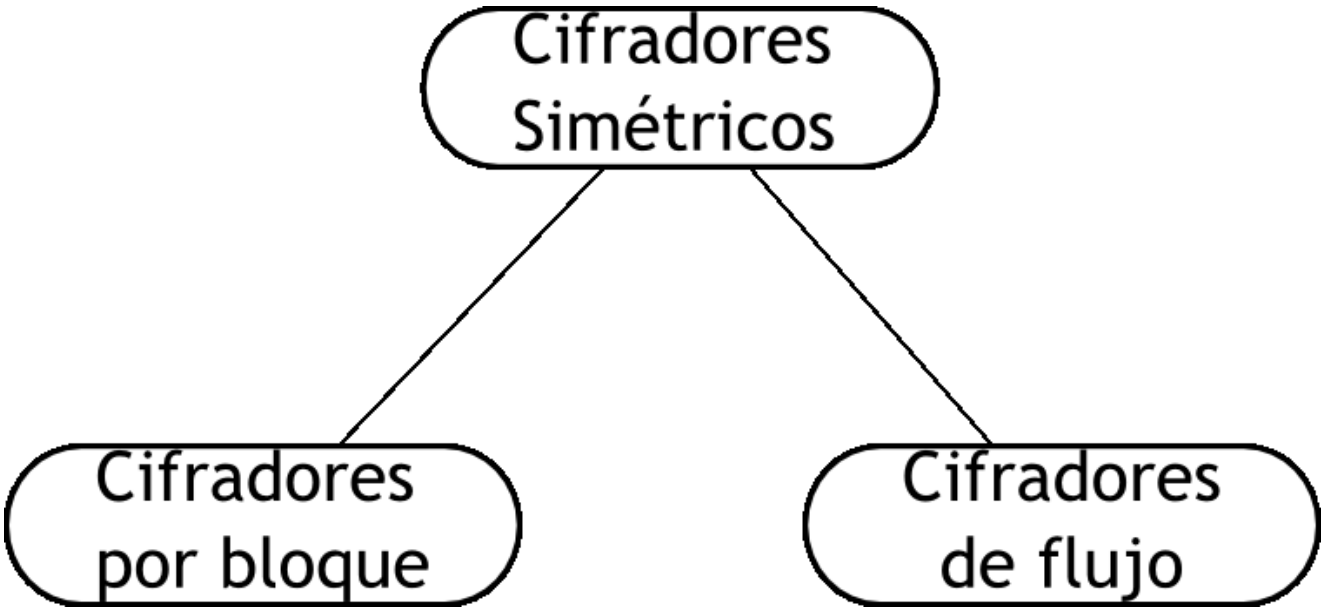
Criptosistema

Un *criptosistema* es una 5-tupla $(\mathcal{P}, \mathcal{C}, \mathcal{K}, \mathcal{E}, \mathcal{D})$, con las siguientes condiciones:

- $\mathcal{P}$ es el conjunto finito de todos los textos en claro posibles
- $\mathcal{C}$ es el conjunto finito de todos los textos cifrados posibles
- $\mathcal{K}$, el *espacio de claves*, es el conjunto finito de todas las *claves* posibles
- $\forall K \in \mathcal{K}$, $\exists E_K \in \mathcal{E}$ (regla de cifrado), $\exists D_K \in \mathcal{D}$ (regla de descifrado)

Cada $E_K : \mathcal{P} \to \mathcal{C}$, $D_K : \mathcal{C} \to \mathcal{P}$, son funciones tal que $\forall x \in \mathcal{P}$, $D_K(E_K(x)) = x$.

Tipos de cifradores simétricos



Cifradores simétricos

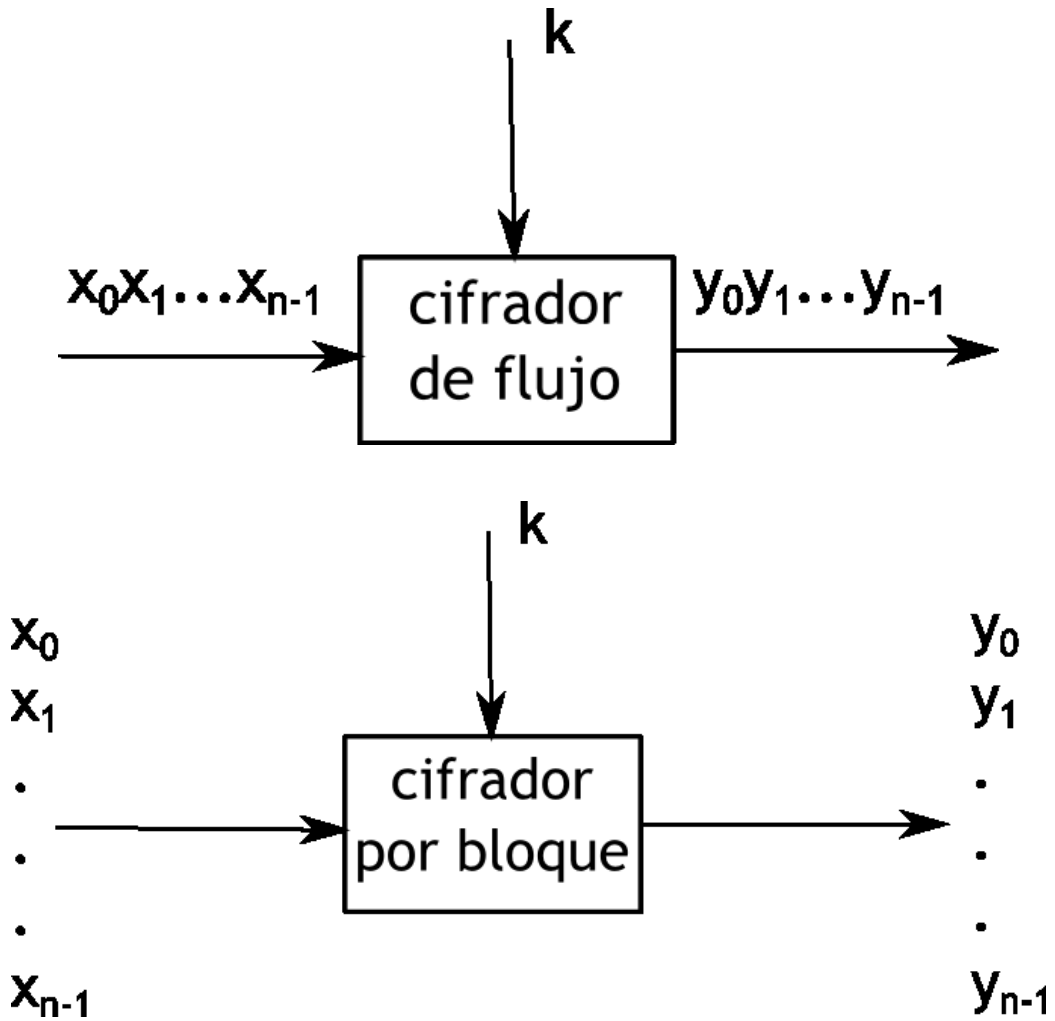
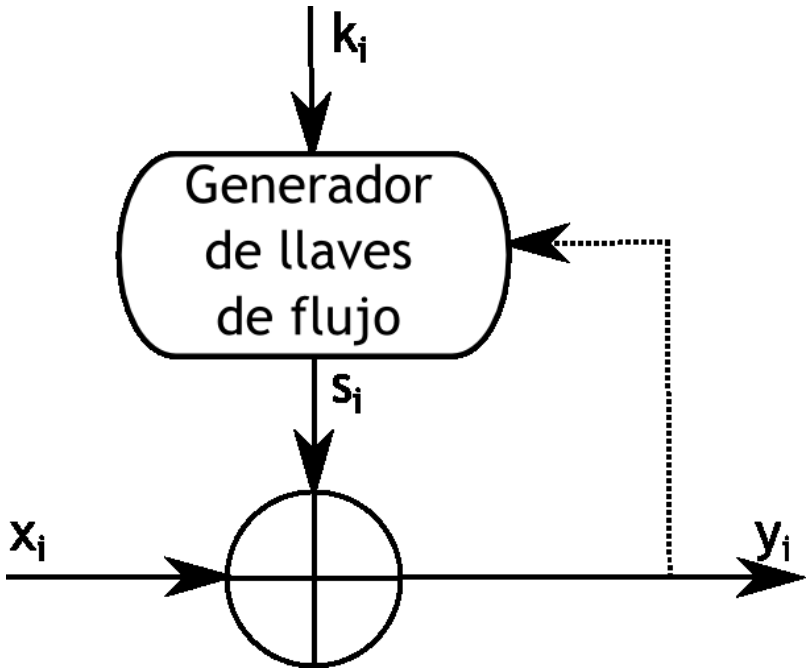


Diagrama cifradores de flujo



Cifradores de flujo.
Cifran bits individualmente. Esto se hace al añadir un bit de un flujo de la llave a un bit del texto en claro.

Los esquemas *síncronos* son en los que la línea punteada en el diagrama no está presente (el cifrado depende exclusivamente de la llave. Son *asíncronos*, cuando existe la dependencia del bit cifrado con los bits cifrados anteriormente (la línea punteada está activa).

Operación

Cifrado y Descifrado de flujo

El texto en claro, el texto cifrado, y el flujo de la llave consiste en bits individuales: $x_i, y_i, s_i \in \{0, 1\}$

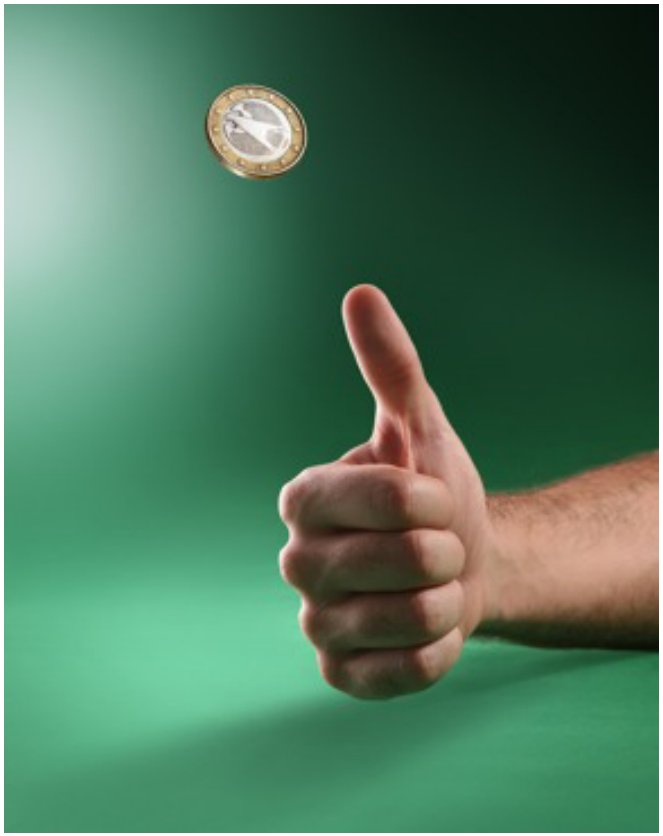
- Cifrado: $y_i = e_{s_i}(x_i) \equiv x_i + s_i \bmod 2$
- Descifrado: $x_i = d_{s_i}(y_i) \equiv y_i + s_i \bmod 2$

Nota: La suma módulo 2 equivale a la operación xor.

Cifradores de Bloque

Cifradores de bloque.
Cifran un bloque completo de bits del texto en claro a la vez con la misma llave. Esto significa que el cifrado de cualquier bit del texto en claro en un bloque dado depende de los otros bits del bloque. En la práctica, la mayoría de los cifradores de bloque esperan bloques de 128 bits (AES), o 64 bits (DES).

Números aleatorios



Generadores de números verdaderamente aleatorios.

- Los generadores de números realmente aleatorios (TRNGs) se caracterizan por el hecho de que su salida no puede ser reproducida. Por ejemplo, si echamos 100 volados y registramos los resultados como una secuencia de bits, dicha secuencia es virtualmente irrepetible (la probabilidad de repetirla es de $1/2^{100}$).
- Los TRNGs están basados en procesos físicos.

Número pseudo-aleatorios

Generadores de números pseudo-aleatorios.

- Los generadores de números pseudo-aleatorios (PRNGs) generan secuencias que pueden ser calculadas a partir de un valor inicial llamado semilla (seed).

Por ejemplo, la función rand(.) del ANSI C es algo así:

$$s_0 = 12345$$
$$s_{i+1} \equiv 1103515245 \cdot s_i + 12345 \bmod 2^{31}, i = 0, 1, \ldots$$

Generadores de números pseudo-aleatorios criptográficamente seguros

Un **generador de números pseudo-aleatorios criptográficamente seguros** (CSPRNGs) es un tipo especial de generador que es impredecible. Dada una secuencia de bits, no existe un algoritmo polinomial que determine el siguiente bit con una probabilidad mayor al 50%. Igualmente, dada una secuencia de bits, es imposible determinar el anterior.

La impredecibilidad de los CSPRNGs es única para la criptografía, por lo que si se toma un generador no diseñado específicamente para criptografía, probablemente no sirva para un producto comercial.

Incondicionalmente seguro

Incondicionalmente seguro.
Un criptosistema es incondicionalmente seguro (o seguro en términos de la teoría de la información) si no puede ser roto aún con recursos informáticos infinitos.

Suponga un criptosistema simétrico con una llave de 10,000 bits que solo pueda ser roto mediante búsqueda exhaustiva (fuerza bruta). Se necesitarían $2^{10,000}$ computadoras. El sistema no es incondicionalmente seguro, pero es computacionalmente seguro (se estima que existen entre 2^{239} y 2^{289} átomos en el universo)

One-time pad

Aquí está un criptosistema incondicionalmente seguro:

One-time pad

Es un cifrador de flujo en el cual:

- el flujo de la llave $s_0, s_1, \ldots$ es generador por un TRNGs
- el flujo de la llave es solamente conocido por los extremos de la comunicación
- cada bit del flujo de la llave s_i es utilizado una única vez.

se le conoce como one-time pad. El one-time pad es incondicionalmente seguro.

Más sobre el one-time pad.

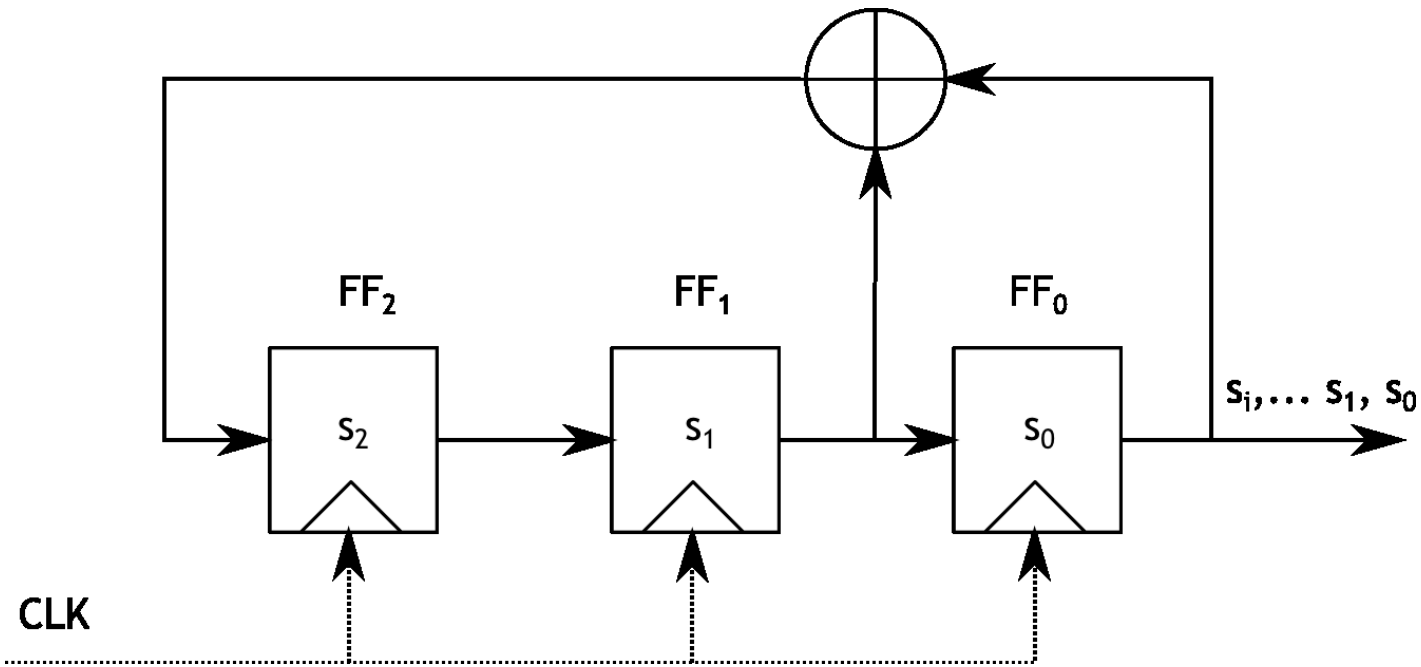
Cada bit del texto cifrado se forma de la siguiente manera:

$$\begin{aligned} y_0 &\equiv x_0 + s_0 \bmod 2 \\ y_1 &\equiv x_1 + s_1 \bmod 2 \\ &\vdots \\ y_{n-1} &\equiv x_{n-1} + s_{n-1} \bmod 2 \end{aligned}$$

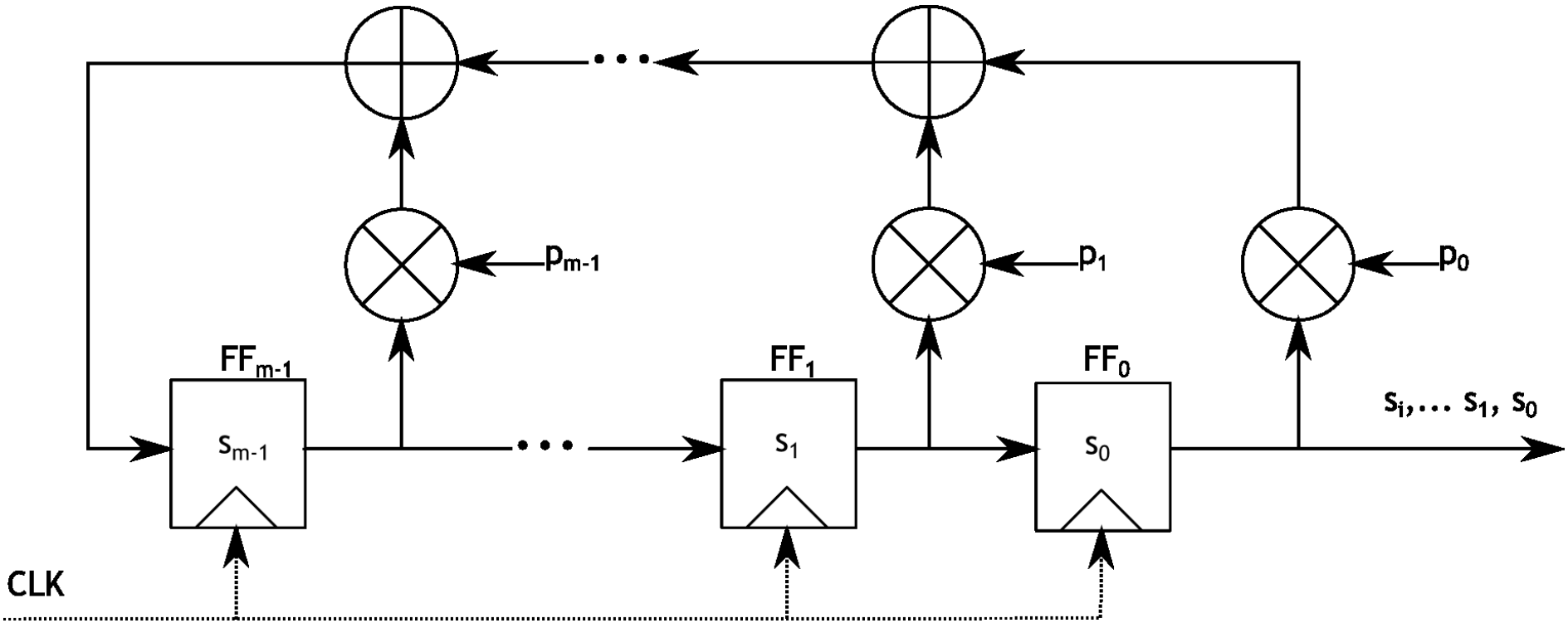
es una ecuación con dos incógnitas por cada bit. Aún si se conoce y_i, los valores de $x_i \in \{0, 1\}$ tienen exactamente la misma probabilidad si se utilizó un TRNG. Sin embargo, los bits aleatorios no se pueden reutilizar, lo que nos lleva al problema de distribución de llaves.

Linear Feedback Shift Register

Un **LFSR** consiste en elementos de almacenamiento sincronizados (flip-flops) y una ruta de retroalimentación. El número de elementos de almacenamiento establece el grado del LFSR. La red de retroalimentación calcula la entrada del último flip-flop como una suma módulo 2 (XOR) de ciertos flip-flops en el registro.



LFSR



DES - Data Encryption Standard

Historia:

- Las compañías financieras necesitaban de un mecanismo de protección que tuviese el visto bueno del gobierno norteamericano.
- El primer llamado para el concurso fue en mayo de 1973, seguido de un segundo llamado en 1974
- No hubo muchas propuestas. IBM presentó Lucifer.
- El gobierno trabajó con la IBM para rediseñar el algoritmo

DES - Data Encryption Standard

Cronograma:

- 1973 - el NBS (El Buró nacional de estándares de los EEUU, ahora NIST) solicita propuestas de criptosistemas para documentos ``no-clasificados''
- 1974 - el NBS repite el requerimiento.
  - IBM responde con una modificación de LUCIFER.
    LUCIFER está basado en una familia de cifradores creados por Horst Feistel a finales de los 1960s.
  - NBS le pide a la NSA (National Security Agency) que lo evalúe.
    Por esas fechas la NSA negaba su propia existencia.
  - La NSA convence a la IBM que reduzca el tamaño estándar de la clave del LUCIFER de 128-bits a 56-bits, esto posibilitaría los ataques por fuerza bruta.
  - . . .

DES

- . . .
  - LA NSA escogió algunos de los parámetros del sistema del cifrador. En particular pidió que se agregara soporte a un tipo de ataques en particular. Estos ataques se descubrirían en los 1990s, y se llamarían ataques diferenciales. Se desconoce si la NSA sabía de su existencia, o si sólo sospechaban.
  - IBM obtiene la patente del DES (Data Encryption Standar).
- 1975 se publican los detalles del algoritmo, excepto algunos criterios de funciones internas. La discusión pública comienza. La gente tenía la incertidumbre si la NSA le había metido mano para su beneficio (introducción de trap-doors).
- . . .

DES

- 1976 se adopta como un estándar para todos los doumentos gubernamentales ``no-clasificados''.
  Data Encryption Standard - FIPS PUB 46.
- Se hace estándar para hardware en 1977
- ANSI X3.92-1981 (hardware + software)
- ANSI X3.106-1983 (modes of operation)
- Australia AS2805.5-1985
- Se reafirmó como estándar hasta 1999 cuando se substituyó por el AES - Advanced Encryption Standard.

Confusión y difusión

De acuerdo con el teorista de la información Claude Shannon, hay dos operaciones primitivas con las cuales los algoritmos de cifrado fuerte pueden ser construídos:

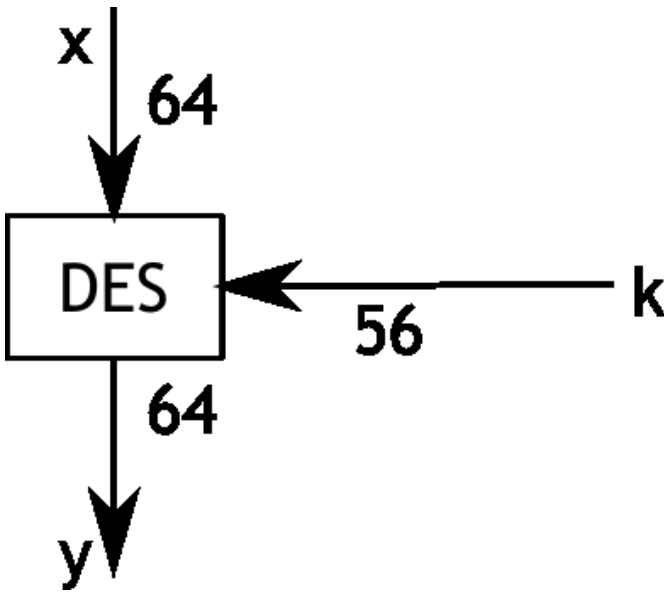
- **Confusión**. Operación de cifrado en la que la relación entre la llave y el texto cifrado se ocultan.
- **Difusión**. Operación de cifrado en la que la influencia de un símbolo del texto en claro se reparte sobre muchos del texto cifrado para ocultar las propiedades estadísticas del texto en claro; i.e., permutaciones y mezcla columnas.

Confusión y difusión

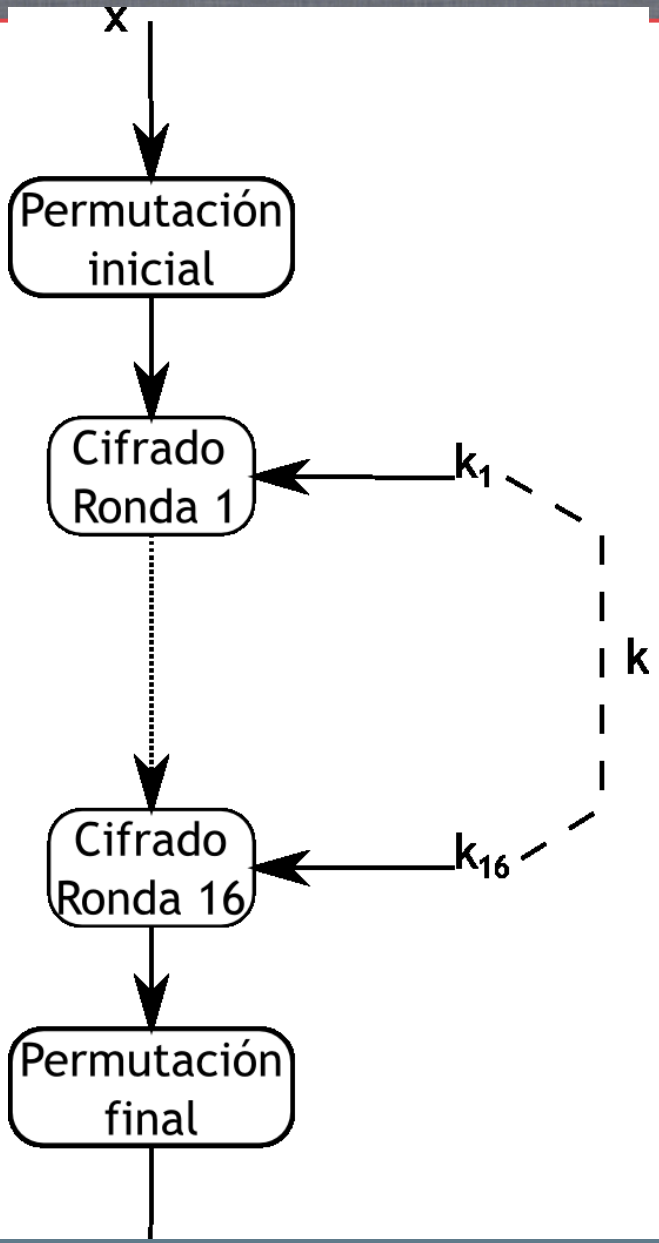
- Utilizar sólo confusión o sólo difusión resulta en cifradores que no son seguros.
- A través de la concatenación de dichas primitivas se construyen cifradores fuertes; estos se conocen como cifradores de producto.
- Todos los cifradores modernos y seguros son cifradores de producto.
- Una secuencia de operaciones de confusión y difusión se le conoce como ronda.
- Las rondas se repiten n veces sobre el texto en claro o el resultante de la ronda anterior. Esto provoca un efecto avalancha.

DES caja negra

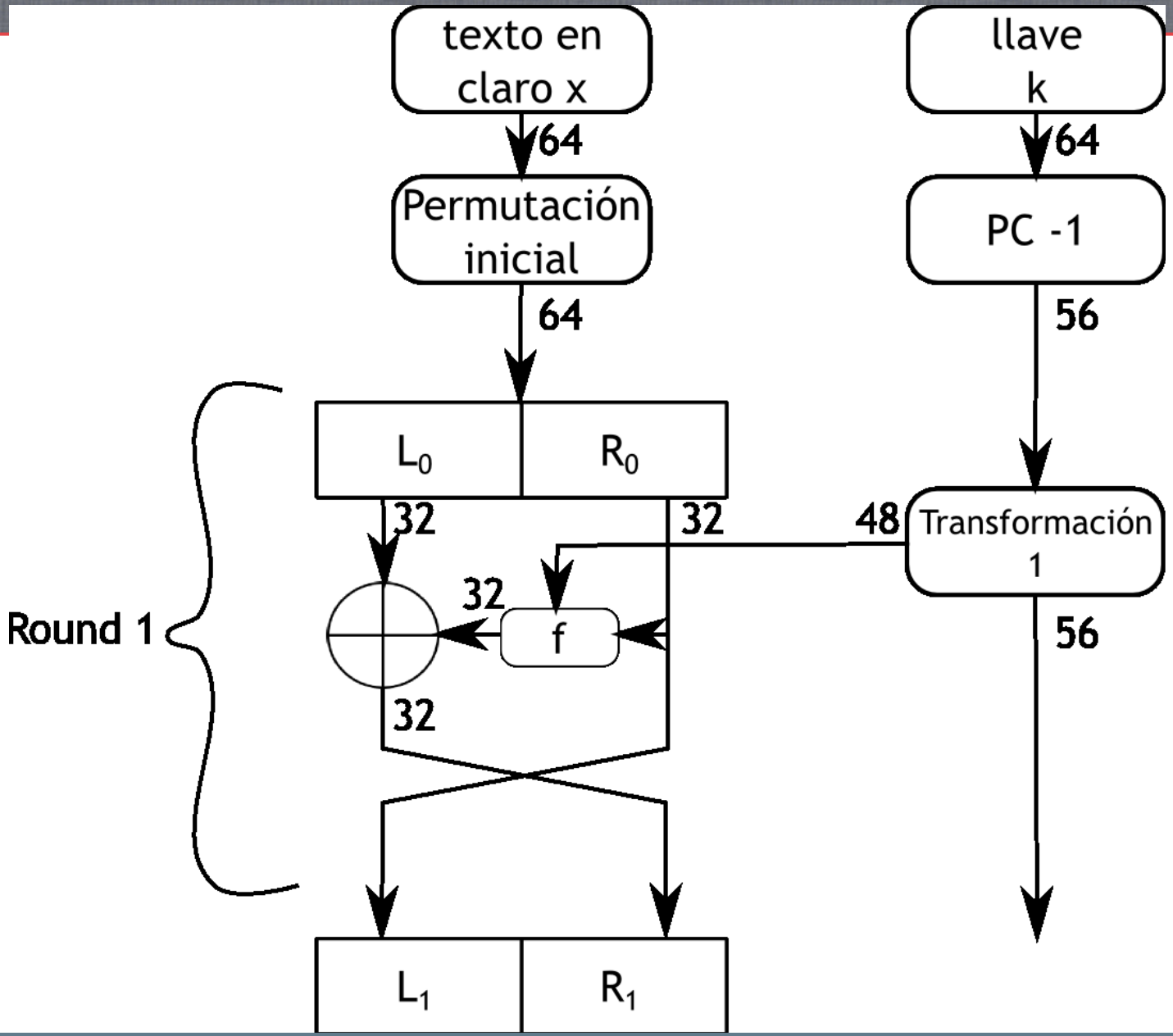
El DES es un cifrador que toma bloques de 64-bits de longitud con una llave de 56-bits.



DES estructura general



DES estructura Feistel



Ataques

- 1977 - Diffie y Hellman sugirieron un diseño de un chip VLSI que puede probar 10^6 llaves/seg. Un equipo con 10^6 de estos circuitos podría romper la clave en cuestión de 10 horas. Costo: USD $20'000,000.00
- 1990 - Eli Biham y Adi Shamir sugirieron un criptoanálisis diferencial
- 1993 - Mitsuru Masui sugirió un criptoanálisis linear

Ataque de la paradoja del cumpleaños en el logaritmo discreto

Dado un primo p, y α y β enteros que no son cero módulo p, es imposible encontrar una x tal que $\alpha^x \equiv \beta \bmod p$, si p es lo suficientemente grande.

Sin embargo, con el ataque del cumpleaños:

- Haga dos listas de longitud $\approx p^{1/2}$
- La primera lista contendrá todos los valores de $\alpha^k \bmod p$, para $\approx p^{1/2}$ valores aleatorios de k.
- La segunda lista contendrá los números $\beta\alpha^{-\ell} \bmod p$, para $\approx p^{1/2}$ valores aleatorios de ℓ.

Cuando haya alguna coincidencia, tenemos que $\alpha^k \equiv \beta\alpha^{-\ell} \bmod p$, por lo que $\alpha^{k+\ell} \equiv \beta \bmod p$. El valor buscado es $x \equiv k + 1 \bmod p - 1 \ldots$

Ataque Meet-in-the-middle

Asuma que Eva capturó un mensaje m y un texto doblemente cifrado $c = E_{k_2}(E_{k_1}(m))$. Calcue y almacene $E_k(m)$ y $D_k(c)$ para todos los valores posibles de la llave k. Compare ambas listas. Debe de existir una coincidencia.

Dado el reducido tamaño de la clave (56-bits):

- En 1998, la Electronic Frontier Foundation rompió una clave en 56 horas: 1536 chips probando 88×10^9 llaves/segundo, con un costo menos a los USD $250,000.
- En 1999, Distribute.Net, en conjunto con la EFF, consiguieron 100,000 equipos voluntarios en internet, y rompieron una clave en 22 horas y 15 minutos.

Suponiendo que usara DES (que no lo usa), cada cuándo cambian su contraseña de Gmail?

Modos de operación

Hemos dicho que DES trabaja con bloques de 64-bits, ¿qué pasa si requerimos cifrar más de 64 bits? Agarramos de 64 bits en 64 bits? (no)

Modos de bloques

- ECB - Electronic Codebook Block
- CBC - Cipher Block Chaining

Modos de flujo

- CFB - Cipher Feedback
- OFB - Output Feedback

Modos de operación

Hemos dicho que DES trabaja con bloques de 64-bits, ¿qué pasa si requerimos cifrar más de 64 bits? Agarramos de 64 bits en 64 bits? (no)

Modos de bloques

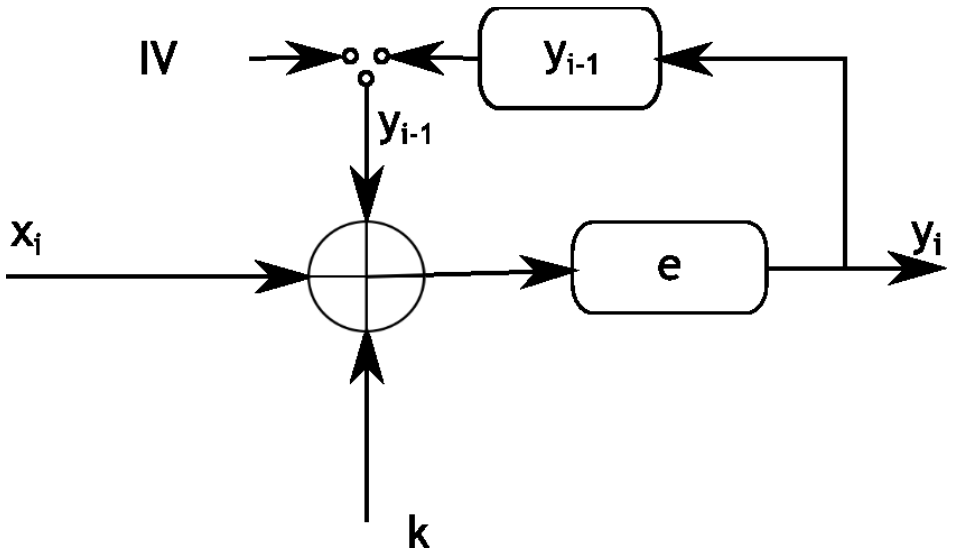
- ECB - Electronic Codebook Block
- CBC - Cipher Block Chaining

Modos de flujo

- CFB - Cipher Feedback
- OFB - Output Feedback

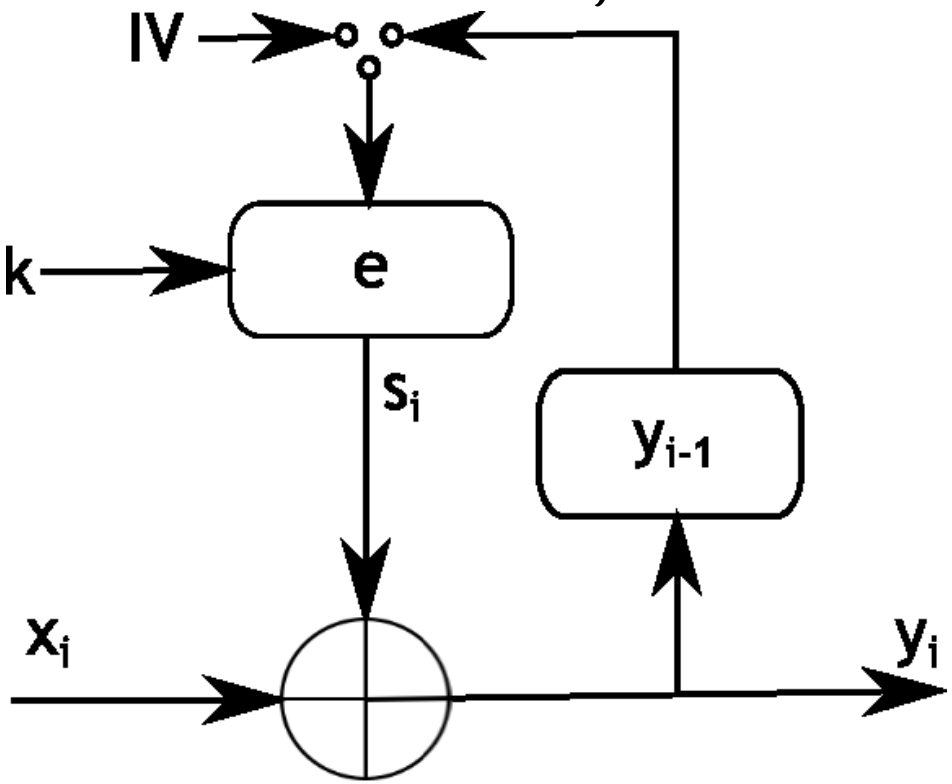
Por bloques

- ECB - El mensaje se rompe en bloques de 64-bits (se rellena con ceros).
- CBC - Hace un xor de la salida anterior con el bloque a cifrar (requiere vector de inicialización).

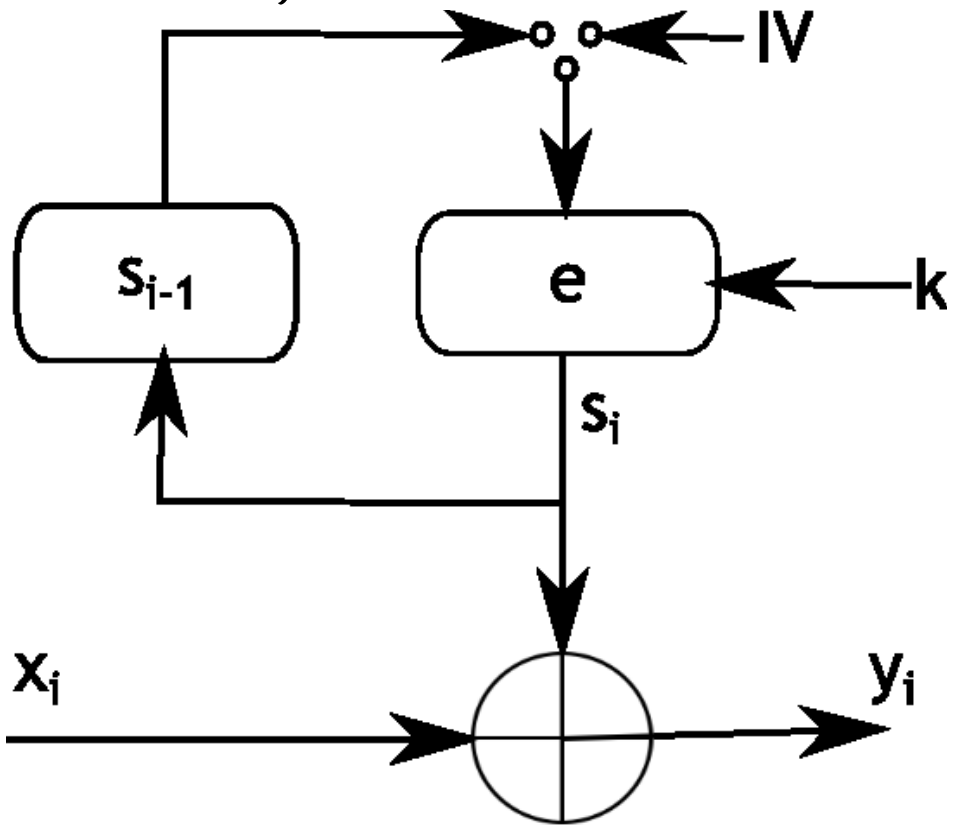


Por flujo

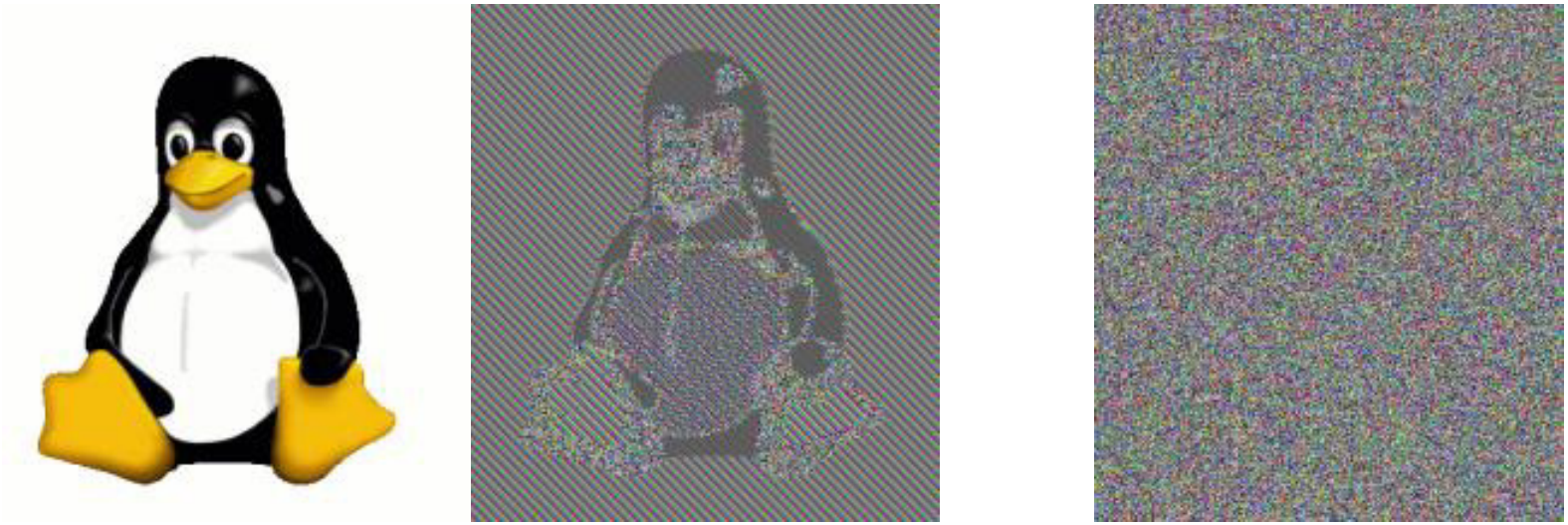
CFB - Se hace un xor de la salida anterior con el mensaje



OFB - El feedback es independiente del mensaje actual



Comparativa

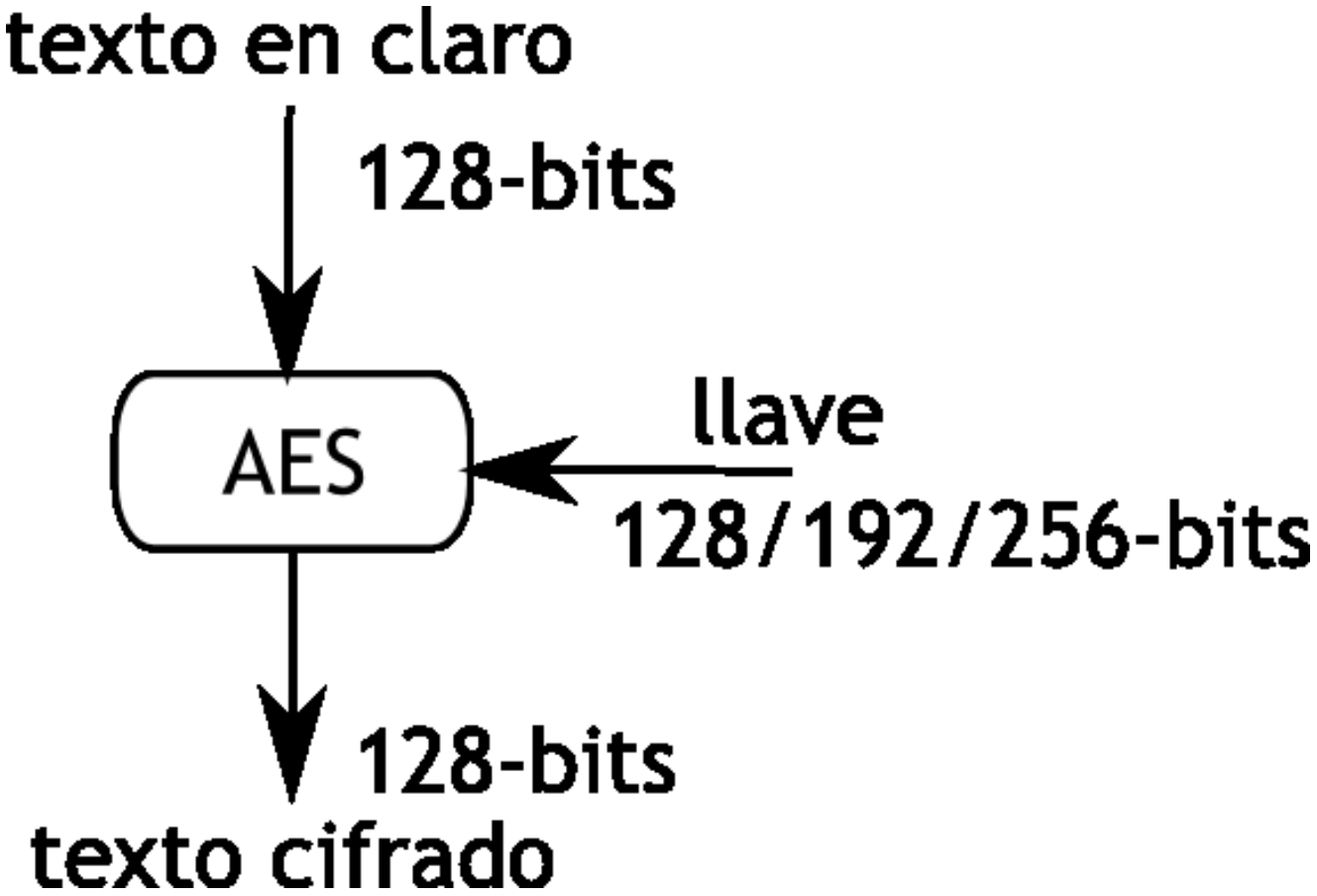


AES - Advanced Encryption Standard

¿Porqué un estándar nuevo?

- Ataques de fuerza bruta
- La solución (Triple DES) lo hace el triple de lento
- DES es eficiente solo para hardware
- Nuevos tipos de ataques
- Utilizar bloques de 64-bits no es útil para todos los escenarios

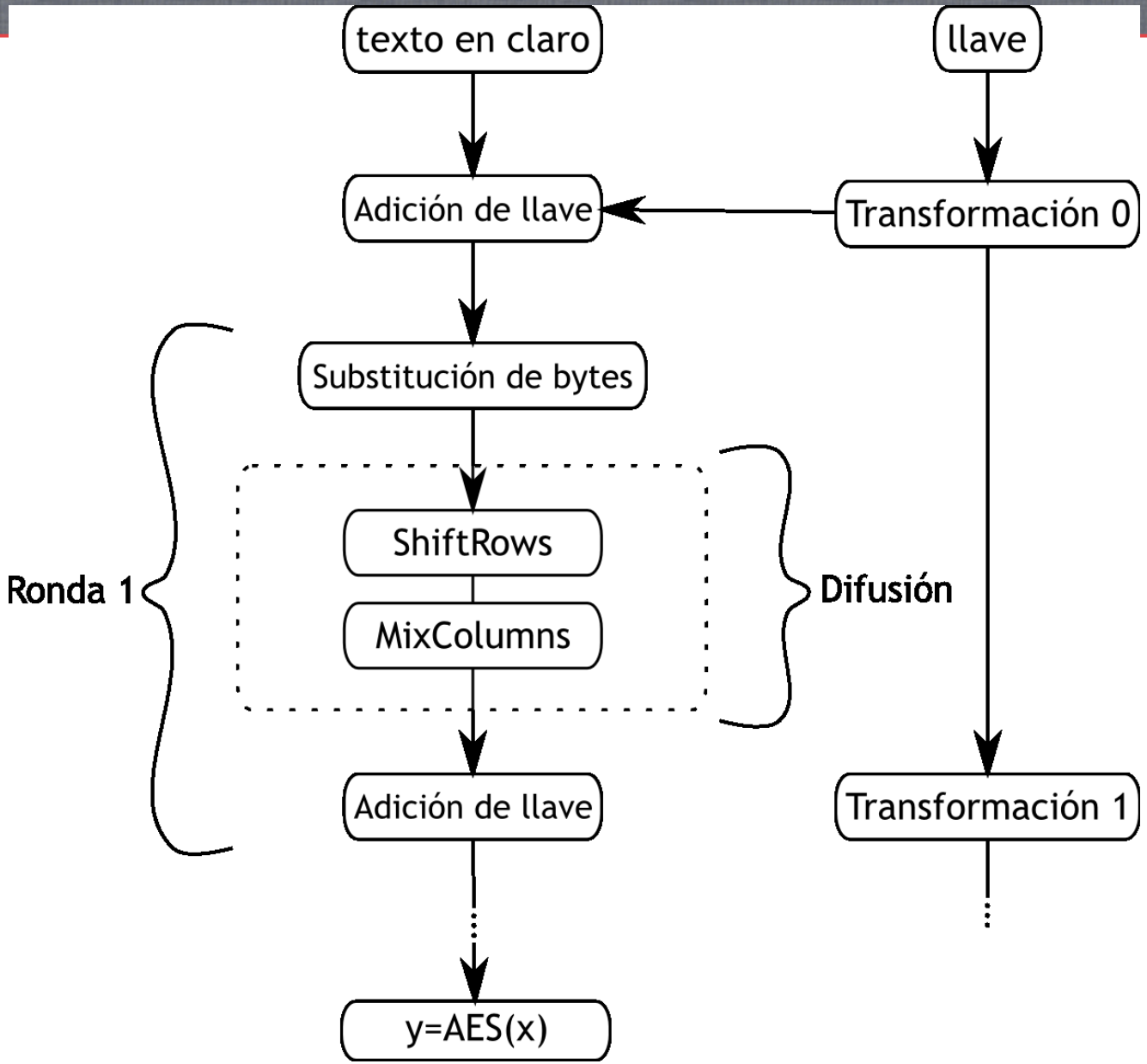
AES - Caja negra



AES - descripción

- AES no utiliza una función Feistel, se desea cifrar todo un bloque por ronda
- Se necesitan 10, 12 o 14 rondas para cifrar con claves de 128, 192 o 256 bits
- En cada ronda hay 3 capas: Adición de llave, de Sustitución de bytes, y de difusión
- la capa de difusión se subdivide en: ShiftRow, que permuta datos a nivel de byte; y MixColumn, que mezcla bloques de 4-bytes dentro de una matriz

AES - rondas


texto en claro
llave
Adición de llave
Transformación 0
Substitución de bytes
ShiftRows
MixColumns
Ronda 1
Difusión
Adición de llave
Transformación 1
y=AES(x)

- Fin de la unidad 6, 7