Criptografía de llave privada en la práctica

Introducción a la Criptografía y a la Seguridad de la Información

Iván Castellanos

Departamento de ingeniería de sistemas e industrial
Universidad Nacional de Colombia

24 de septiembre de 2019

Llaves aleatorias

- La seguridad de muchos sistemas de cifrado depende de la generación de números (llaves) aleatorios grandes
- En caso que hayan llaves muy probables tendremos una vulnerabilidad en el sistema de cifrado
- *Aleatorio* en terminos de información significa que *no puede ser comprimido*

Ejemplo

En el DES si seleccionaramos de manera totalmente aleatoria una llave de 56 bits k, tendriamos un espacio de 2^{56} llaves y necesitariamos intentar en promedio 2^{55} llaves para encontrar k. Si $k = f(s)$ donde s es de 16 bits en promedio necesitariamos solo 2^{15} intentos

Llaves aleatorias

Definición (Generador aleatorio de bits)

Un genrador aleatorio de bits es un dispositivo o algoritmo que produce una secuencia de bits estadisticamente *independientes* de manera *imparcial* (probabilidad uniforme)

Nota

Podemos utilizar un generador aleatorio de bits para generar enteros de manera aleatoria. Para generar un entero en el rango [0, n] se puede generar una secuencia de tamaño $\lfloor \log n \rfloor + 1$

Llaves aleatorias

Definición (Secuencia aleatoria)

Una secuencia s es aleatoria si y solo si $K(s) = |s| + C$ donde $K(s)$ es la mínima longitud de la posible descripción de s

Teorema

Para una secuencia s es teoréticamente imposible calcular $K(s)$

Nota

Esto quiere decir que es imposible de manera *determinística* hacer un generador alteatorio

Generadores pseudoaleatorios

Definición (PRBG)

Un *generador pseudoaleatorio de bits (pseudorandom bit generator)* es un algoritmo determinístico que con una entrada aleatoria (*semilla*) de tamaño k genera una secuencia binaria de tamaño l, $l \gg k$

Nota

La idea es a partir de una pequeña secuencia realmente aleatoria crear una secuencia larga que para un atacante no tenga diferencia con una realmente aleatoria

Ejemplo

Dado una semilla x_0 podemos tener la recurrencia lineal $x_n = ax_{n-1} + b \bmod m$ donde $a, b, m \in \mathbb{Z}$. Este generador es *predecible* como para ser utilizado en criptografía

Generadores pseudoaleatorios

Nota

Al utilizar un PRBG el tamaño k de la semilla tiene que ser lo suficientemente grande para que una busqueda sobre todas las semillas sea infactible para un adversario

Definición

Test del siguiente bit Un PRBG pasa el *test del siguiente bit* si no hay algoritmo polinomial que dados los primeros l bits de una secuencia s, pueda predecir el $l+1$-ésimo bit de s con una probabilidad significamente mayor a $\frac{1}{2}$

Definición (CSPRBG)

Un PRBG que pasa el test del siguiente bit es llamado un *generador pseudoaleatorio de bit criptográficamente seguro*

Semillas aleatorias

Definición (Eventos físicamente aleatorios)

Existen algunos eventos físicos que son aleatorios y se peuden utilizar como la semilla de los PRBG

Ejemplo

- Tiempo entre tecleos en un computador
- Contenido en buffers I/O
- Ruido térmico de un diodo semiconductor
- Inestabilidad de frecuencia de un oscilador de corrido libre

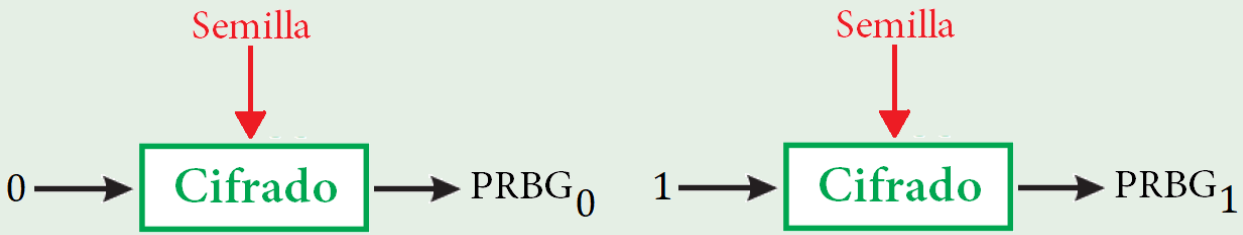
Nota

Combinando valores de varios de estos eventos podemos obtener semillas aleatorias resistentes a ataques

PRBG

Ejemplo

Podemos utilizar un sistema de cifrado con alta difusión y confusión para realizar un PRBG



Nota

Podriamos utilizar el DES o el AES para construir un PRBG

Nota

Luego de varias iteraciones podemos utilizar una salida del PRBG como una nueva semilla

Test estadísticos

- Es infactible realizar de manera completa el test del siguiente bit para probar si una la pseudoaleatoridad de una secuencia
- Podemos utilizar *test estadísticos* (polinomiales) para probar, con cierta probabilidad, la no pseudoaleatoriedad de una secuencia
- Si una secuencia pasa varios de estos test estadísticos tenemos cierta confianza sobre la pseudoaleatoridad de una secuencia

Ejemplo

Sean F_1 y F_2 test estadisticos, y m una secuencia de bits, $F_1(m)$ falla con una probabilidad de $= 0.2$ y $F_2(m)$ falla con una probabilidad de $= 0.999$, podemos decir entonces que m no es pseudoaleatorio con una alta probabilidad a causa de F_2

Test estadísticos

Definición (Hipótesis estadística)

Una *hipótesis estadística* H_0 es una afirmacion sobre una distribución de variables aleatorias

Definición (Prueba de hipótesis estadística)

Una *prueba de hipótesis estadística* es un procedimiento en el cual a partir de la observación de muestras de variables aleatorias se *acepta* o se *rechaza* H_0, este test no es defintivo sino que es *probabilístico*

Definición (Nivel de significancia)

El *nivel de significancia* α de una hipótesis estadística H_0 es la probabilidad de rechazar H_0 cuando H_0 es verdadero

Test estadísticos

- Para la prueba de la aleatoridad de nuestros PRBG H_0 va a ser la hipótesis que sado una secuencia s esta fue generada por un generador aleatorio de bits
- Si el nivel se significancia α es muy alto podriamos rechazar secuencias que fueron generadas aleatoreamente (error tipo I)
- Si α es muy bajo podemos acpetar secuencia que no fueron generadas aleatoreamente (error tipo II)

Definición (Test estadístico)

Un *test estadístico* corresponde al cálculo (en tiempo polinomial) de un *estadístico muestral* $X(s)$ a partir de una muestra s, el cual se aproxima a una distribución de probabilidad y con el que podremos hacer una prueba de hipótesis estadística

Test estadísticos

Ejemplo

Sea X un estadístico muestral con $X \sim \chi^2$ con 5 grados de libertad. Para lograr un nivel de significancia $\alpha = 0{,}025$ (2.5 %) se calcula el *umbral* x_α tal que $P(X > x_\alpha) = \alpha$, en este caso $x_\alpha = 12{,}8325$. Si para la secuencia s, $X(s) > x_\alpha$ entonces s falla el test estadístico, en caso contrario s lo pasa

Ejemplo

Sea X un estadístico muestral con $X \sim N(0,1)$. Para lograr un nivel de significancia $\alpha = 0{,}05$ (5 %) se calcula x_α tal que $P(X > x_\alpha) = P(X < -x_\alpha) = \frac{\alpha}{2}$, en este caso $x_\alpha = 1{,}96$. Si para la secuencia s, $X(s) > x_\alpha$ o $X(s) < -x_\alpha$ entonces s falla el test estadístico, en caso contrario s lo pasa

Test de frecuencia

- Una secuencia de bits aleatoria tiene una probabilidad muy baja de generar solo 1's o 0's
- En una secuencia de bits aleatoria se espera que la cantidad de 0's y de 1's sea aproximadamente la misma

Definición (Test de frecuencia)

Sean n_0 y n_1 el numero de 0's y 1's de la s respectivamente, definimos el estadístico del *test de frecuencia (monobit test)* como $X_1(s) = \frac{(n_0 - n_1)^2}{n}$, que se aproxima a una distribución χ^2 con 1 grado de libertad

Test de frecuencia

- Una secuencia de bits aleatoria tiene una probabilidad muy baja de generar solo 1's o 0's
- En una secuencia de bits aleatoria se espera que la cantidad de 0's y de 1's sea aproximadamente la misma

Definición (Test de frecuencia)

Sean n_0 y n_1 el numero de 0's y 1's de la s respectivamente, definimos el estadístico del *test de frecuencia (monobit test)* como $X_1(s) = \frac{(n_0 - n_1)^2}{n}$, que se aproxima a una distribución χ^2 con 1 grado de libertad

Ejercicio

Calcule el estadístico del test de frecuencia de la secuencia (en hexadecimal) $s = e3114ef249$

Test serial

- En una secuencia de bits aleatoria se espera que la cantidad de ocurrencias de 00, 01, 10 y de 11 a lo largo de la secuencia sea aproximadamente la misma

Definición (Test serial)

Sean n_{00}, n_{01}, n_{10} y n_{11} las ocurrencias de 00, 01, 10 y 11 en s respectivamente, definimos el estadístico del *test serial (two-bit test)* como $X_2(s) = \frac{4}{n-1}(n_{00}^2 + n_{01}^2 + n_{10}^2 + n_{11}^2) - \frac{2}{n}(n_0^2 + n_1^2) + 1$, que se aproxima a una distribución χ^2 con 2 grados de libertad

Test de poker

- En una secuencia de bits aleatoria partida en cadenas de tamaño m, se espera que la cantidad de ocurrencias de cada posible cadena de tamaño m a lo largo de la secuencia sea aproximadamente la misma

Definición (Test de poker)

Sea $m \in \mathbb{Z}^+$ tal que $\lfloor \frac{n}{m} \rfloor \geq 5 * (2^m)$ y sea $k = \lfloor \frac{n}{m} \rfloor$. Divida la secuencia en k partes de tamaño m que no se sobrelapen (una partición). Sea n_i el numero se ocurrencias del tipo i de tamaño m, $i = 0, ..., 2^m - 1$, definimos el estadístico del *test de poker* como $X_3(s) = \frac{2^m}{k}(\sum_{i=0}^{2^m-1} n_i^2) - k$, que se aproxima a una distribución χ^2 con $2^m - 1$ grados de libertad

Test de corrido

Definición (Corrido)

Un *corrido* es una secuencia de bits idénticos

- En una secuencia de bits aleatoria no se espera que hayan muchos corridos largos

Definición (Test de corrido)

El valor esperado de corridos de tamaño i en una secuencia aleatoria de tamaño n es $e_i = \frac{n-i+3}{2^{i+2}}$. Sea $k = \text{máx}\{i \in \mathbb{Z}^+ \mid e_i \geq 5\}$ y sean B_i y G_i el numero de corridos de 1's y 0's de tamaño i en s respectivamente para $i = 1, ..., k$, definimos el estadístico del *test de corrido* como $X_4(s) = \sum_{i=1}^{k}(\frac{(B_i-e_i)^2}{e_i}) + \sum_{i=1}^{k}(\frac{(G_i-e_i)^2}{e_i})$, que se aproxima a una distribución χ^2 con $2k-2$ grados de libertad

Test de autocorrelación

- En una secuencia de bits aleatoria debería haber una baja correlación entre la secuencia y la secuencia corrida d bits

Definición (Test de autocorrelación)

Sea $d \in \mathbb{Z}$, $1 \leq d \leq \lfloor \frac{n}{2} \rfloor$, el número de bits iguales en el d-corrimiento de s es $A(d) = \sum_{i=0}^{n-d-1} s_i \oplus s_{i+d}$. Definimos el estadístico del *test de autocorrelación* como $X_5(s) = 2\frac{A(d) - \frac{n-d}{2}}{\sqrt{n-d}}$, que se aproxima a una distribución $N(0,1)$. Note que tanto valores pequeños como valores de $A(d)$ son inesperados

Ejemplo test estadísticos

Ejemplo

Considere la secuencia no aleatoria s de tamaño $n = 160$ obtenida repitiendo la secuencia e3114ef249 (en hexadecimal) 4 veces

1. Frecuencia: $n_0 = 84$, $n_1 = 76$, luego $X_1(s) = 0{,}4$
2. Serial: $n_{00} = 44$, $n_{01} = 40$, $n_{10} = 40$, $n_{11} = 35$, luego $X_2(s) = 0{,}6252$
3. Poker: para $m = 3$ y $k = 53$, los bloques 000, 001, 010, 011, 100, 101, 110 y 111 aparecen 5, 10, 6, 4, 12, 3, 6 y 7 veces respectivamente, luego $X_3(s) = 9{,}6415$
4. Corridos: $e_1 = 20{,}25$, $e_2 = 10{,}0625$, $e_3 = 5$, luego $k = 3$. Hay 25, 4 y 5 corridos de 1's de tamaños 1, 2 y 3 respectivamente y 8, 20 y 12 corridos de 0's de tamaños 1, 2 y 3 respectivamente, luego $X_4(s) = 31{,}7913$
5. Autocorrelación: para $d = 8$ $A(d) = 100$, luego $X_5(s) = 3{,}8933$

Con un nivel de significancia de $\alpha = 0{,}05$ los umbrales para X_1, X_2, X_3, X_4 y X_5 son 3.8415, 5.9915, 14.0671, 9.4877 y 1.96 respectivamente, luego s pasa los tests de frecuencia, serial, y poker pero falla el de corridos y el de autocorrelación

Nota

Los valores de umbrales para distribuciones χ^2 y $N(0,1)$ se pueden encontrar precalculados en tablas

Modos de operación

Nota

Dos bloques de 128 bits (16 ASCII o 4 UTF-32) iguales cifrados con el AES usando la misma llave nos genera el mismo texto cifrado, para mensajes largos esto es suceptible a repeticiones

Definición (Modos de operación)

Los modos de operación son los procedimiento que permiten utilizar de manera repetida y segura un cifrado por bloques (como el AES) con una misma llave

Nota

Con los modos de operación tendremos algo similar a la generación de llaves a partir de una semilla

Electronic Codebook Mode

Definición (ECB)

El electronic codebook mode (ECB) mapea para cada $i \in \{0,1,...,2^j-1\}$ (en AES $j=128$) el valor $E_k(i)$, luego para cada $m = m_0 m_1 ... m_{n-1}$, $E_k(m_i) = c_i$ para $\forall i \in \{0,1,...,n-1\}$, Finalmente $c = c_0 c_1 ... c_{n-1}$

Nota

En resumen partimos un mensaje en bloques y cada bloque es cifrado directamente con la llave que tenemos

Electronic Codebook Mode

Usando ECB se tienen varios problemas

- Se conoce la longitud del mensaje, pues es igual a la del texto cifrado
- Es suceptible a repetición de bloques
- Un atacante (activo) puede mover bloques o reemplazar bloques y el descifrado puede ser valido con los bloques en diferente orden

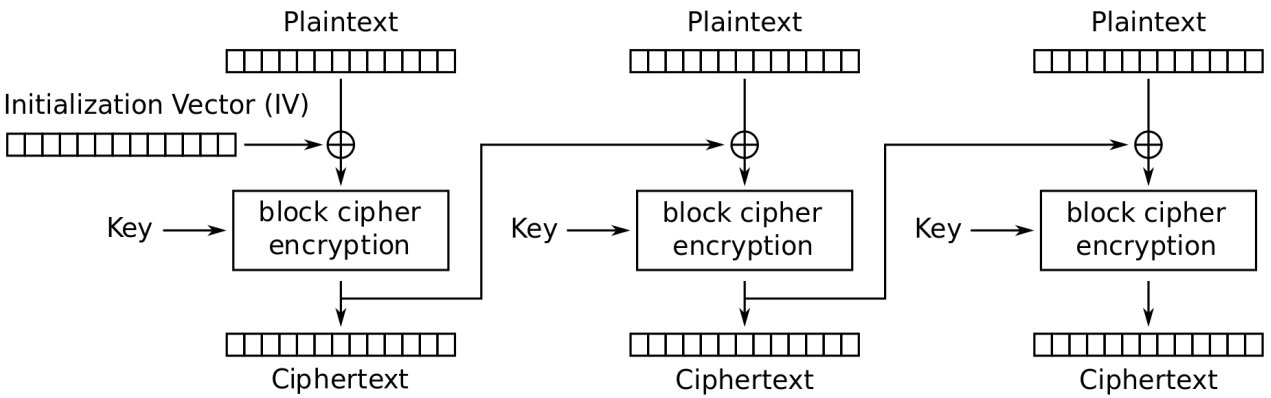
Nota

El primer problema lo tienen todos los cifrados en bloque que no hagan *padding*

Cipher Block Chaining Mode

Definición (CBC)

Sea $m = m_0 m_1 ... m_{n-1}$, el *cipher block chaining (CBC) mode* cifra de la siguiente manera $c_0 = E_k(IV \oplus m_0)$ y $c_i = E_k(c_{i-1} \oplus m_i)$, donde IV es un vector de inicialización de tamaño b que puede ser público



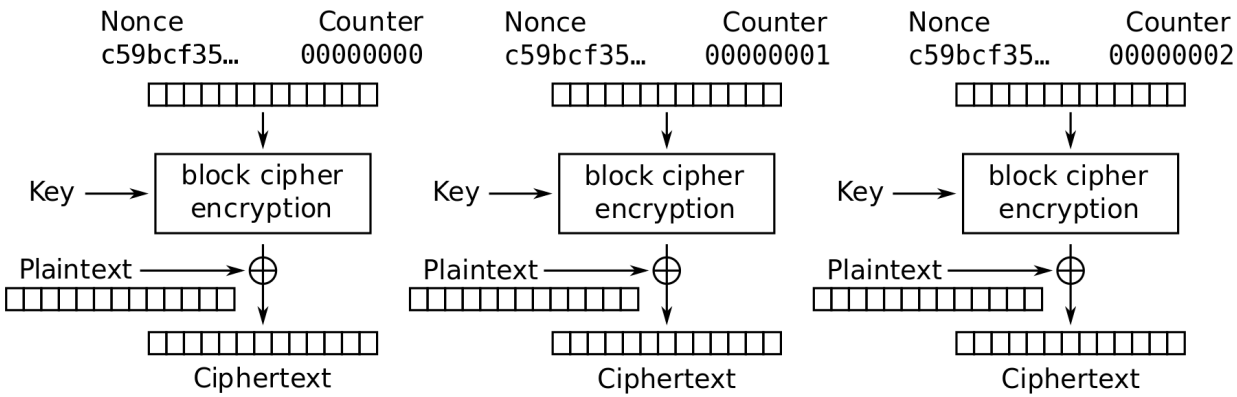
Nota

La idea del modo CBC es que el texto cifrado de un bloque anterior impacte en el siguiente. Es costoso dado que no es paralelizable

Counter Mode

Definición (CTR)

Sea $m = m_0 m_1 ... m_{n-1}$, el *counter (CTR) mode* cifra de la siguiente manera $c_i = E_k(nonce|i) \oplus m_i$, donde *nonce* es un vector de inicialización que puede ser público



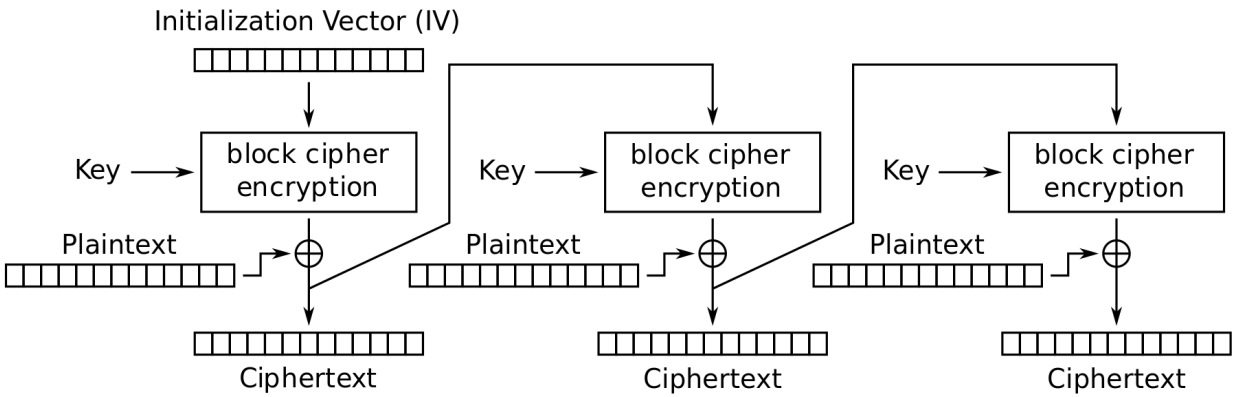
Nota

Usando el AES *nonce* podría ser de 64 bits y el contador de otros 64 bits. Este caso es análogo a utilizar el OTP con un PRBG basado en AES. Cada bloque es maleable

Cipher Feedback Mode

Definición (CFB)

Sea $m = m_0 m_1 ... m_{n-1}$, el *cipher feedback (CFB) mode* cifra de la siguiente manera $c_i = E_k(x_i) \oplus m_i$, $x_0 = IV$ y $x_i = c_{i-1}$, donde IV es un vector de inicialización que puede ser público



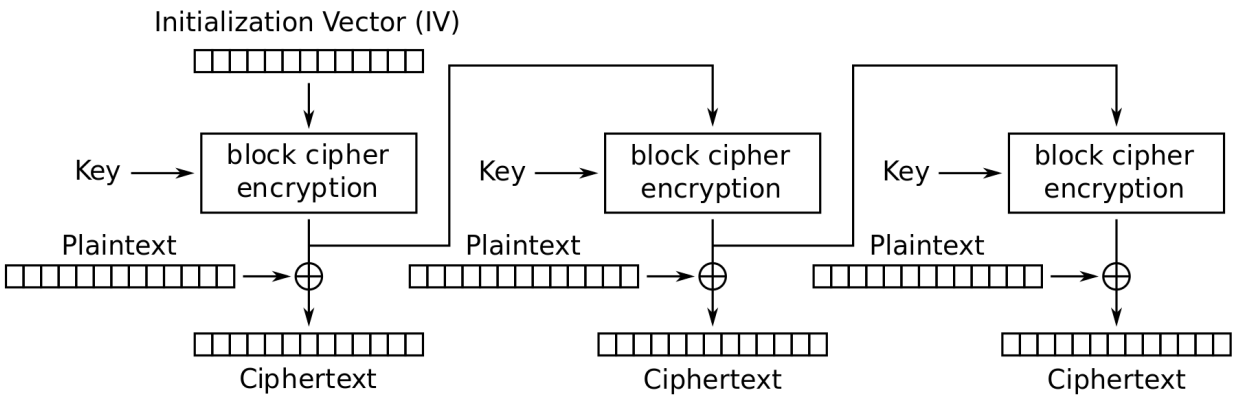
Nota

El mayor problema del CFB es que si hay algun error en el envío de algún bloque de cifrado se daña todo el descifrado

Output Feedback Mode

Definición (OFB)

Sea $m = m_0 m_1 ... m_{n-1}$, el *output feedback (OFB) mode* cifra de la siguiente manera $c_i = E_k(x_i) \oplus m_i$, $x_0 = IV$ y $x_i = E_k(x_{i-1})$, donde IV es un vector de inicialización que puede ser público



Nota

El OFB con AES se puede ver como un OTP con un PRBG basado en AES, donde se uitiliza un vector inicial para el calculo de las llaves