Generador Binario Pseudoaleatorio Basado en la Combinación de Registros de Desplazamiento con Retroalimentación Lineal, mediante Funciones por Mayoría

Andrés Francisco Farías – Andrés Alejandro Farías

Departamento Académico de Ciencias Físicas, Matemáticas y Naturales
Universidad Nacional de La Rioja, La Rioja. Argentina
(afarias665@yahoo.com.ar, andres_af86@hotmail.com)

Abstract
El presente documento expone el procedimiento de construcción de un Generador binario pseudoaleatorio basado en la combinación de registros de desplazamiento con retroalimentación lineal (Linear Feedback Shift Register, LFSR). El proceso incluye la descripción del modelo, la estructura de cada generador, selección de las funciones booleanas que cuenten con las mejores propiedades criptográficas, la definición de la combinación final. Por último, para verificar la aleatoriedad de las secuencias obtenidas, se aplican a las mismas un conjunto de pruebas estadísticas de aleatoriedad.

Keywords: LFSR, cipher, key, Boolean function, non-linearity

1 Introducción

Un requisito fundamental de ese tipo de generadores se relaciona con la calidad de la secuencia generada. Entre otras características se exige imprevisibilidad y facilidad de implementación, pero, fundamentalmente un período con una longitud significativa.

Es en esos términos que se propone un modelo que responda a tales exigencias. La modalidad elegida se basa en la combinación no lineal de secuencias producidas por cuatro LFSR [1], [2].

El procedimiento de construcción de un generador pseudoaleatorio de ese estilo requiere de varias etapas:

- Definición esquemática del modelo.
- Función por mayoría.
- Elección de los distintos LFSR.
- Selección de funciones booleanas de cuatro variables en base a sus propiedades criptográficas.
- Conformación del generador con los componentes ya seleccionados.
- Clave y el procedimiento para generar los estados iniciales de los LFSR.

- Elección de las pruebas estadísticas a utilizar y los criterios de análisis de los resultados.
- Puesta en funcionamiento y realización de las pruebas de aleatoriedad necesarias sobre las secuencias obtenidas.

2 Definición esquemática del modelo

El generador propuesto en este trabajo, está conformado por cuatro LFSR, que tienen cada uno, dos funciones booleanas de filtrado no lineal, que producen secuencias binarias, las que luego son combinadas mediante funciones mayoría. Las secuencias obtenidas alimentan dos funciones booleanas de cuatro variables, cuyos resultados son sometidos a una operación XOR, según la figura 1:

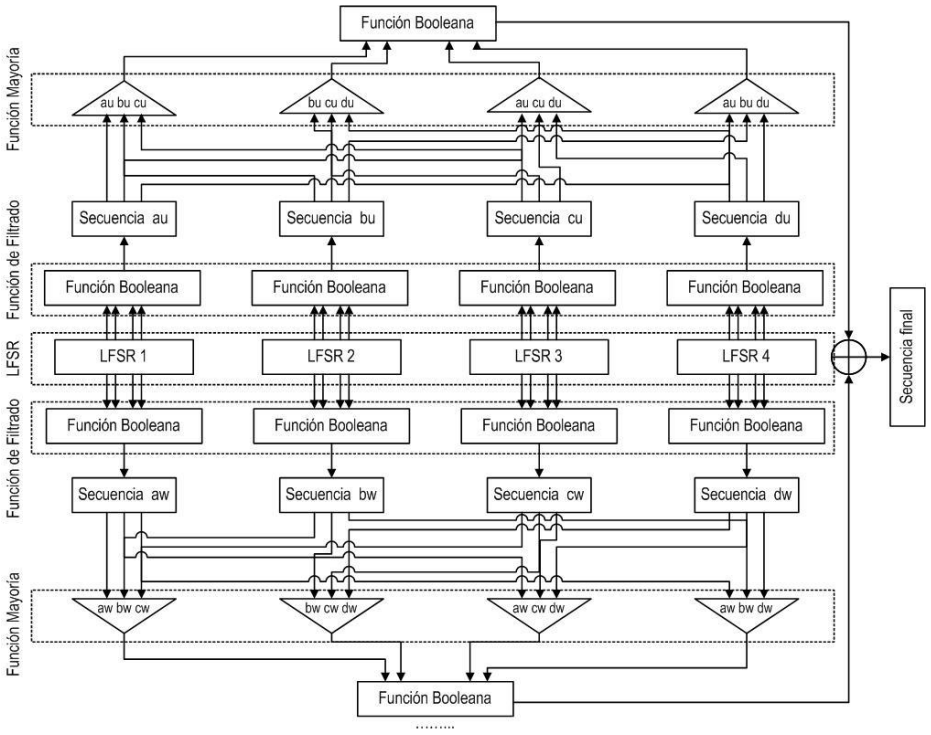

Fig. 1. Esquema generador binario pseudoaleatorio

3 Función por mayoría

En el esquema se indica la combinación de secuencias mediante función por mayoría, el número de secuencias debe ser impar y el valor binario es el que más se repite:

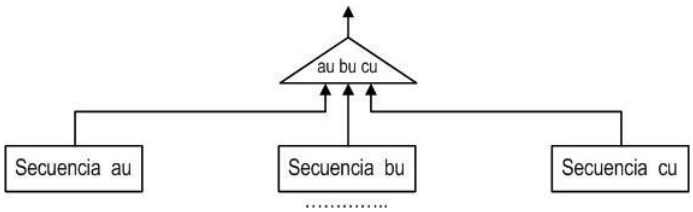

Fig. 2. Función por mayoría

4 Elección de los distintos LFSR

Las longitudes y polinomios primitivos de cada LFSR, que componen el generador, son las siguientes [3], [4], [5].

Tabla 1. LFSR, longitudes y polinomios primitivos del Generador

LFSR	Longitud	Polinomios primitivos
1	47	$P(x) = x^{47} + x^{32} + x^{24} + x^{11} + 1$
2	61	$P(x) = x^{61} + x^{57} + x^{26} + x^{3} + 1$
3	59	$P(x) = x^{59} + x^{54} + x^{46} + x^{26} + 1$
4	53	$P(x) = x^{53} + x^{50} + x^{41} + x^{20} + 1$

5 Selección de las funciones booleanas

5.1 Propiedades criptográficas deseables .

A continuación se indican algunas de las propiedades criptográficamente más significativas, adoptadas para este trabajo [6], [7], [8].

- **Función Balanceada**: Una función booleana de n-variables $\boldsymbol{f}$ es balanceada si $\boldsymbol{w(f) = 2n - 1}$. Esta propiedad es deseable para evitar ataques criptodiferenciales. La función es balanceada cuando el primer coeficiente del espectro de Walsh-Hadamard, es igual a cero: $\boldsymbol{F(0) = 0}$.
- **No Linealidad**: Valores altos de esta propiedad reducen el efecto de los ataques por criptoanálisis lineal. La No Linealidad de una función booleana puede ser calculada con la transformada de Walsh-Hadamard, $NL_f = \frac{1}{2} \cdot (2^n - |WH_{max}(f)|)$
- **Grado Algebraico**: El grado algebraico de una función, es el número de entradas más grande que aparece en cualquier producto de la Forma Normal Algebraica. Es deseable que sean valores altos.
- **SAC**: El Criterio de Avalancha Estricto requiere los efectos avalancha de todos los bits de entrada. Una función booleana se dice que satisface SAC sí y solo sí, la Ecuación 3, es balanceada para toda u con w(u)=1, $f(x) \oplus f(x \oplus u)$

Siguiendo los criterios arriba indicados las funciones booleanas aceptadas, son:

Tabla 2. Funciones de cuatro variables adoptadas

f_{NAF}
$f_{5775} = a \oplus b \oplus a \cdot b \oplus a \cdot c \oplus a \cdot d$
$f_{4722} = a \oplus b \oplus a \cdot c \oplus b \cdot c \oplus c \cdot d$
$f_{4529} = a \oplus c \oplus a \cdot c \oplus b \cdot c \oplus c \cdot d$
$f_{4393} = a \oplus c \oplus a \cdot d \oplus b \cdot d \oplus c \cdot d$
$f_{2402} = b \oplus a \cdot c \oplus b \cdot c \oplus d \oplus c \cdot d$
$f_{3981} = a \oplus a \cdot c \oplus b \cdot c \oplus d \oplus c \cdot d$
$f_{3338} = a \oplus a \cdot b \oplus c \oplus b \cdot c \oplus b \cdot d$
$f_{3672} = a \oplus a \cdot b \oplus c \oplus a \cdot c \oplus a \cdot d$
$f_{5911} = a \oplus b \oplus a \cdot b \oplus b \cdot c \oplus b \cdot d$
$f_{5056} = a \oplus b \oplus a \cdot d \oplus b \cdot d \oplus c \cdot d$

6 Conformación del generador combinacional

El generador combinacional queda de la siguiente manera:.

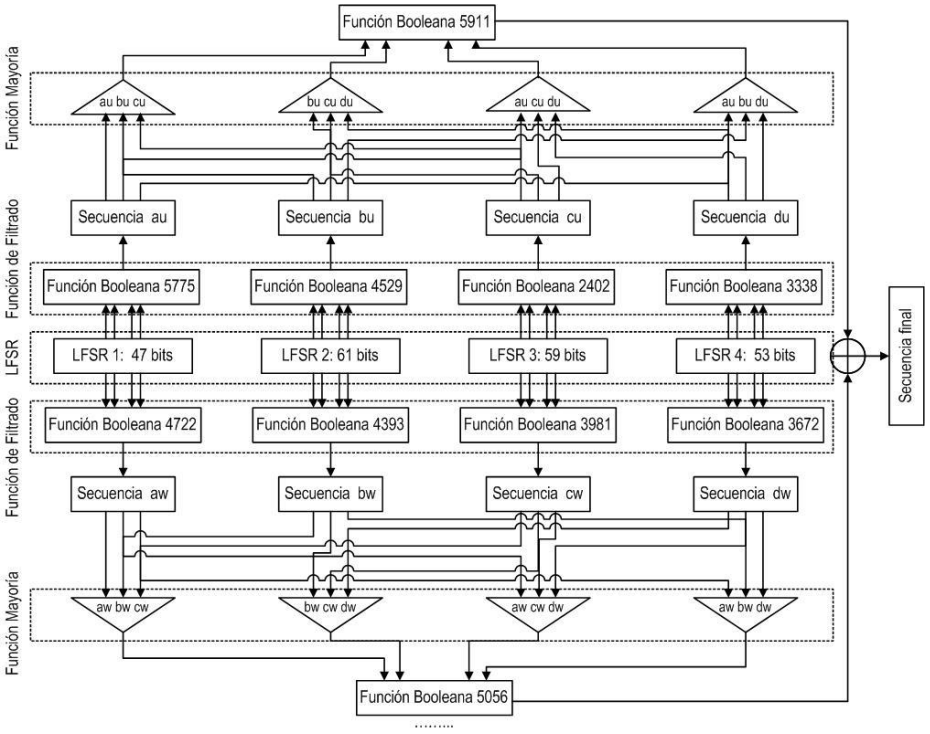

Fig. 3. Generador Combinacional

7 Clave

Para originar los estados iniciales de los distintos LFSR se realiza un proceso que utiliza una clave de 32 caracteres, que expresada en código ASCII (American Standard Code for Information Interchange), tiene longitud de 256 bits.

Se aceptan solamente las letras del alfabeto inglés (minúsculas y mayúsculas) y los números del sistema de numeración decimal, es decir un total de 62 caracteres.

La clave es sometida a un proceso criptográfico, que se indica en la Figura 5.

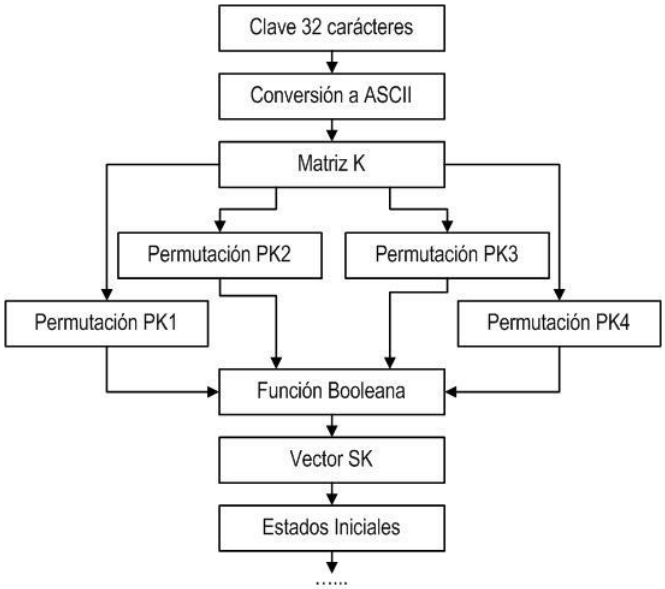

Fig. 4.. Clave para el generador

8 Permutaciones

8.1 Generador congruencial multiplicativo

El generador tiene la siguiente expresión: [9]

$$x_{i+1} = (a_x \cdot x_i)mod\ m_x \qquad (1)$$

Donde: $a_x = multiplicador, m_x = módulo, x_0 = semilla$

Tabla 3. Vectores, módulos, multiplicadores y semillas

Vector	módulo	multiplicador	semilla
PK1	1048576	1747	3249
PK2	1048576	1753	3271
PK3	1048576	1759	3301
PK4	1048576	1777	3347

8.2 Generación de los estados iniciales

La función booleana que procesa los cuatro vectores $PK1, PK2, PK3$ y PK4 es la siguiente: $MK = PK2 \oplus (PK1 \cdot PK3) \oplus (PK2 \cdot PK3) \oplus (PK1 \cdot K4) \oplus (PK2 \cdot PK4)$

De la operación resulta un vector SK[j] de 256 bits, que es el que proveerá los estados iniciales de los LFSR, en forma secuencial.

9 Elección de las pruebas estadísticas

Fueron seleccionadas algunas pruebas de la Norma NIST Special Publication 800-22, del trabajo de Rukhin (et al.) [10].

9.1 Prueba de frecuencia

El propósito de esta prueba es determinar si el número de unos y ceros en una secuencia es aproximadamente el mismo que se espera de una secuencia verdaderamente aleatoria. La prueba evalúa la cercanía de la fracción de unos a ½, que es decir, el número de unos y ceros en una secuencia debe ser aproximadamente el mismo. Todas las pruebas posteriores dependen de la aprobación de esta prueba.

9.2 Prueba de frecuencia en un bloque

La meta de esta prueba es determinar si la frecuencia de unos en un bloque de M bits es aproximadamente $M / 2$, como se esperaría bajo un supuesto de aleatoriedad.

9.3 Prueba de rachas

Una racha de longitud k consta de exactamente k bits idénticos y está acotada antes y después con un poco del valor opuesto. El propósito de la prueba de rachas es determinar si el número de rachas unos y ceros de varias longitudes es lo esperado para una secuencia aleatoria.

9.4 Prueba de rachas de unos en un bloque

El fin de esta prueba es determinar si la longitud de la ejecución más larga de las dentro de la secuencia probada es consistente con la longitud de la serie más larga de las que cabría esperar en una secuencia aleatoria. Tenga en cuenta que una irregularidad en la longitud esperada de la serie más larga implica que también hay una irregularidad en la longitud de la serie más larga de ceros.

9.5 Prueba de sumas acumuladas

Determina si la suma acumulativa de las secuencias parciales que ocurren en la secuencia probada es demasiado grande o demasiado pequeña en relación con el comportamiento esperado de esa suma acumulada para secuencias aleatorias.

9.6 Prueba de entropía aproximada

El enfoque de esta prueba es la frecuencia de todas las posibles superposiciones patrones de m bits en toda la secuencia. El propósito de la prueba es comparar la frecuencia de bloques superpuestos de dos longitudes consecutivas / adyacentes (m, m + 1) contra el resultado esperado para un secuencia aleatoria.

10 Pruebas sobre el generador

Se analizaron cien secuencias binarias, obtenidas del generador a partir de cien claves distintas.

El nivel de significancia adoptado para las pruebas estadísticas es de $\alpha = 0{,}01$. La hipótesis nula es:

$$H_0 \rightarrow p\_valor > 0{,}01$$

Debido al gran volumen de procesamiento requerido, se desarrolló un programa escrito en lenguaje C++, con los algoritmos correspondientes al generador y a las pruebas estadísticas. Es decir que el software calculó las secuencias binarias y simultáneamente realizó las pruebas sobre las mismas.

11 Interpretación de los resultados

Teniendo los resultados se realizan dos procesos para la interpretación de los mismos:

- Proporción de muestras que pasan las pruebas.
- Prueba de Uniformidad de los p-valor
    - Tabla de frecuencia e histograma
    - Prueba de Bondad de Ajuste

Se aplica la prueba de Bondad de Ajuste χ^2 aplicando la siguiente expresión:

$$\chi^2 = \sum_{i=1}^{10} \frac{\left(F_i - \frac{s}{10}\right)^2}{\frac{s}{10}} \qquad (2)$$

Donde: $F_i = \text{Frecuencia de la clase i}$ $s = \text{Cantidad de muestras}$

El primer procedimiento se realiza considerando los resultados de todas las pruebas y el segundo se realiza en forma individual. En todos los casos se deben superar todas las pruebas para aceptar los resultados.

11.1 Proporción de muestras que pasan las pruebas

Para el análisis de los resultados, se determina la proporción de muestras que superan las pruebas, y con esos datos se construye un gráfico de puntos, luego se verifica si los mismos caen dentro de los límites superior e inferior, donde k es el número de muestras.

$$LS, LI = (1-\alpha) \pm 3 \cdot \sqrt{\alpha \cdot (1-\alpha)/k} \qquad (3)$$

En nuestro caso k = 100 y el nivel de significancia elegido es: α = 0.01, los límites quedan:. $LS = 1{,}02$ y $LI = 0{,}96$

Se consideran todas pruebas, los resultados se indican en la tabla

Tabla 4. Pruebas

Pruebas	Proporción	Superior	Inferior
Frecuencias	0,98	1,02	0,96
Frecuencias en un Bloque	1,00	1,02	0,96
Rachas	1,00	1,02	0,96
Rachas de Unos en un Bloque	0,97	1,02	0,96
Sumas Acumuladas Adelante	0,98	1,02	0,96
Sumas Acumuladas Atrás	0,99	1,02	0,96
Entropía Aproximada	0,98	1,02	0,96

En el gráfico se aprecia el resultado, en definitiva la secuencia que entrega el generador supera las pruebas de aleatoriedad.

:

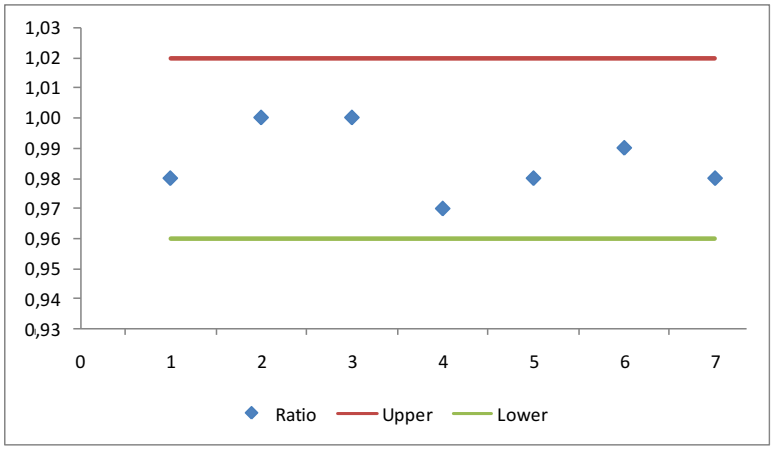

Fig. 5. Gráfico de puntos

11.2 Prueba de bondad de ajuste

Este control se ejecuta para cada prueba sobre las cien muestras, con los resultados de las frecuencias de p-valor obtenidos.

Tabla 5. Pruebas χ^2

Pruebas	χ^2	$\chi^2 ref$	Pasa
Frecuencias	0,658	0,0001	Sí
Frecuencias en un Bloque	0,172	0,0001	Sí

Rachas	0,679	0,0001	Sí
Rachas de Unos en un Bloque	0,817	0,0001	Sí
Sumas Acumuladas Adelante	0,043	0,0001	Sí
Sumas Acumuladas Atrás	0,720	0,0001	Sí
Entropía Aproximada	0,834	0,0001	Sí

11.3 Análisis final

En base a los resultados de las pruebas se realiza una tabla resumen.

Tabla 6. Análisis final

Análisis	Pruebas	Resultados
Proporción de secuencias que pasan las pruebas	Todas	Supera
Distribución uniforme de p-valor	Frecuencias	Supera
	Frecuencias dentro de un bloque	Supera
	Rachas	Supera
	La más larga racha de unos en un bloque	Supera
	Sumas acumuladas adelante	Supera
	Sumas acumuladas atrás	Supera
	Entropía estimada	Supera

En definitiva las secuencias que entrega el generador son pseudoaleatorias.

12 Conclusiones

La generación de bits aleatorios de alta calidad criptográfica resulta de alto interés, en consecuencia, se desarrolló un generador de secuencias binarias pseudoaleatorias de elevado período y complejidad lineal. Para ello se implementó un dispositivo que combina mediante función por mayoría, secuencias producidas por LFSR que sufren un filtrado no lineal con el auxilio de funciones booleanas

Los LFSR que componen cada generador tienen polinomios de conexión primitivos, lo que asegura un elevado período en la secuencia resultante.

La función booleana es la responsable del proceso no lineal, asegura las mejores prestaciones criptográficas, partiendo de criterios tales como ser balanceadas y tener alta no linealidad.

Realizado el proceso de selección, las funciones, las mismas fueron incorporadas al generador, que luego se puso en funcionamiento para generar las secuencias respectivas y con distintos valores de claves.

Los resultados fueron sometidos a un conjunto de pruebas de aleatoriedad, que mostraron valores positivos, por lo que el modelo propuesto se considera válido para la generación de secuencias pseudoaleatorias de buena calidad criptográfica.

13 Referencias

[1] Massodi, F., Alam, S. and Bokhari, M., "A Analysis of Linear Feedback Shift Registers in Stream Ciphers", International Journal of Computer Application, 16 (17), pp. 0975 – 887, 2012.

[2] Menezes, A.,Van Oorschot, P. and Vanstone, S., "Handbook of Applied Cryptography", Massachusetts Institute of Technology, 1996.

[3] Parr, C. and Pelzl, L., Understanding Cryptography, Springer, 2010.

[4] Stahnke, W., "Primitive Binary Polynomials", Mathematics of Computation, 27. 124, pp. 977-980, 1973.

[5] Seroussi, G., "Table of Low-Weight Binary Irreducible Polynomials", Computer Systems Laboratory, 1998.

[6] Clark, J., Jacob, J., Maitra, S., Stanica, P.: Almost Boolean Functions: The Design of Boolean Functions by Spectral Inversion. Computational intelligence. 20. (3), 450—462 (2004)

[7] Braeken, A.: Cryptographic Properties of Boolean Functions and S-Boxes. Faculteit Ingenieurswetenschappen. Katholieke Universiteit Leuven (2003)

[8] Elhosary, A., Hamdy, N., Farag, I., Rohiem, I.: State of the Art in Boolean Functions Cryptographic Assessment. International Journal of Computer Networks and Communications Security.1. (3), 88--94 (2013)

[9] Fishman, G.: Multiplicative Congruential Random Number Generators with Modulus 2ß : An Exhaustive Analysis for ß = 32 and a Partial Analysis for ß = 48. Mathematics of Computation. 54. (189), 33--344 (1990)

[10] Rukhin, A., Soto, J., Nechvatal, J., Smid, M., Barker, E., Leigh, S., Levenson, M., Vangel, M., Banks, D., Heckert, A., Dray, J., and Vo, S., "A Statistical Prueba Suite for Random and Pseudorandom Number Generators for Cryptographic Applications", National Institute of Standards and Technology, (2000).