Cryptography And Cryptanalysis

Ph. D. Course/ 2019-2020

Introduced By

Dr. Faez Hassan Ali

Lecture Four-2

Basic Efficiency Criteria of LFSR's-Systems

Linear complexity (LC) Criterion

Definition: The **linear complexity** of an infinite binary sequence S, denoted LC(S), is defined as follows:

(i). if S is the zero sequence S = 0, 0, 0, . . . , then LC(S) = 0;

(ii). if no LFSR generates S, then LC(S) = ∞;

(iii). Otherwise, LC(S) is the length of the shortest LFSR that generates S.

Definition: The LC of a finite binary sequence S^n, denoted LC(S^n), is the length of the shortest LFSR that generates a sequence having S^n as its first n terms.

Remark (properties of linear complexity) Let S and T be binary sequences.

(i). For any $n \geq 1$, the LC of the subsequence S^n satisfies $0 \leq LC(S^n) \leq n$.

(ii). $LC(S^n) = 0$ if and only if S^n is the zero sequence of length n.

(iii). $LC(S^n) = n$ if and only if $S^n = 0, 0, 0, \ldots, 0, 1$.

(iv). If S is periodic with period N, then $LC(S) \leq N$.

(v). $LC(S \oplus T) \leq LC(S) + LC(T)$.

Linear complexity (LC) Criterion

Linear Complexity Profile

Definition: If $S^n = s_0, s_1, \ldots, s_{n-1}$ is a finite binary sequence, the sequence $LC_1, LC_2, \ldots, LC_n$ is called the LC profile of S^n.

The LC profile of a sequence S can be graphed by plotting the points (N, LC_N), $N \geq 1$, in the N×LC plane and joining successive points by a horizontal line followed by a vertical line, if necessary. The graph of a LC profile is non-decreasing. Moreover, a (vertical) jump in the graph can only occur from below the line LC=N/2; if a jump occurs, then it is symmetric about this line. It's important to show that the expected LC of a random sequence should closely follow the line LC=N/2.

Example: Consider

S^{20}=1,0,0,1,0,0,1,1,1,1,0,0,0,1,0,0,1,1,1,0.

LC profile of S is 1, 1, 1, 3, 3, 3, 3, 5, 5, 5, 6, 6, 6, 8, 8, 8, 9, 9, 10, 10, 11, 11, 11, 11, 14, 14, 14, 14, 15, 15, 15, 17, 17, 17, 18, 18, 19, 19, 19, 19,....

Figure (1) shows the graph of the LC profile of S.

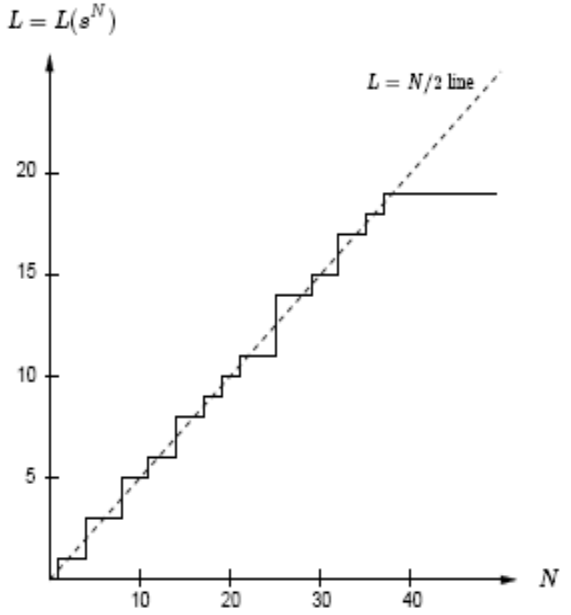

Figure (1): Linear complexity profile of the 20-periodic sequence.

Example (H.W.): the LC profile of the sequence S defined as:

$$s_i = \begin{cases} 1, & \text{if } i = 2^j - 1 \text{ for some } j \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

the line LC=N/2 as closely as possible. That is, $LC(S^N) = \lfloor (N+1)/2 \rfloor$ for all $N \geq 1$. However, the sequence S is clearly non-random.

Linear complexity (LC) Criterion

Berlekamp-Massey Algorithm

Berlekamp-Massey algorithm is an efficient algorithm for determining the LC of a finite binary sequence S^n of length n. The algorithm takes n iterations, with the N^{th} iteration computing the LC of the subsequence S^N consisting of the first N terms of S^n.

Definition: Consider the finite binary sequence $S^{N+1}=s_0,s_1,\ldots,s_{N-1}, s_N$. For $C(D)=1+c_1D+\ldots+c_rD^r$, let $\langle r,C(D)\rangle$ be an LFSR that generates the subsequence $S^N =s_0, s_1,\ldots, s_{N-1}$. The next discrepancy d_N is the difference between s_N and the $(N+1)^{st}$ term generated by the LFSR:

$$d_N=(s_N+\sum_{i=1}^{r} c_i s_{N-i}) \bmod 2$$

Linear complexity (LC) Criterion

Berlekamp-Massey algorithm

INPUT: a binary sequence $S^n = s_0, s_1, s_2, \ldots, s_{n-1}$ of length n.

OUTPUT: the linear complexity $L(S^n)$ of S^n, $0 \leq L(S^n) \leq n$.

PROCESS: 1. Initialization. $C(D) \leftarrow 1$, $r \leftarrow 0$, $m \leftarrow -1$, $B(D) \leftarrow 1$, $N \leftarrow 0$.

2. While $(N < n)$ do the following:

2.1 Compute the next discrepancy d. $d \leftarrow (s_N + \sum_{i=1}^{L} c_i s_{N-i}) \bmod 2$.

2.2 If $d = 1$ then do the following:

$T(D) \leftarrow C(D)$, $C(D) \leftarrow C(D) + B(D).D^{N-m}$.

If $r \leq N/2$ then $r \leftarrow N + 1 - r$, $m \leftarrow N$, $B(D) \leftarrow T(D)$.

2.3 $N \leftarrow N + 1$.

3. Return(r).

Remark: At the end of each iteration of step 2, <r,C(D)> is an LFSR of smallest length which generates S^N. Hence, Berlekamp-Massey algorithm can also be used to compute the LC profile of a finite sequence.

Linear complexity (LC) Criterion

Example Table shows the steps of Berlekamp-Massey algorithm for computing the LC of the binary sequence S^n = 0, 0, 1, 1, 0, 1, 1, 1, 0 of length n=9. This sequence is found to have LC 5, and an LFSR which generates it is $\langle 5, 1+ D^3 + D^5 \rangle$.

Remark: Let S^n be a finite binary sequence of length n, and let the LC of S^n be LC. Then there is a unique LFSR of length LC which generates S^n if and only if LC ≤ n/2.

s_N	d	T(D)	C(D)	r	M	B(D)	N
-	-	-	1	0	-1	1	0
0	0	-	1	0	-1	1	1
0	0	-	1	0	-1	1	2
1	1	1	$1+D^3$	3	2	1	3
1	1	$1+D^3$	$1+D+D^3$	3	2	1	4
0	1	$1+D+D^3$	$1+D+D^2+D^3$	3	2	1	5
1	1	$1+D+D^2+D^3$	$1+D+D^2$	3	2	1	6
1	0	$1+D+D^2+D^3$	$1+D+D^2$	3	2	1	7
1	1	$1+D+D^2$	$1+D+D^2+D^5$	5	7	$1+D+D^2$	8
0	1	$1+D+D^2+D^5$	$1+D^3+D^5$	5	7	$1+D+D^2$	9

Linear complexity (LC) Criterion

Definition : product of m distinct variables is called an m^{th} order product of the variables. Every Boolean function $f(x_1,x_2,...,x_n)$ can be written as a modulo 2 sum of distinct m^{th} order products of its variables, $0 \le m \le n$; this expression is called the algebraic normal form of f. The nonlinear order of f is the maximum of the order of the terms appearing in its algebraic normal form.

Example: the Boolean function $f(x_1,x_2,x_3,x_4,x_5)=1 \oplus x_2 \oplus x_3 \oplus x_4x_5 \oplus x_1x_3x_4x_5$ has nonlinear order 4.

Remark: Suppose that n maximum-length LFSRs, whose lengths $r_1,r_2,...,r_n$ are pairwise distinct and greater than 2, are combined by a nonlinear function $f(x_1,x_2,...,x_n)$ which is expressed in algebraic normal form. Then the LC of the keystream is $f(r_1,r_2,...,r_n)$. (The expression $f(r_1,r_2,...,r_n)$ is evaluated over the integers rather than over Z_2).

Let CF=Fn, so that in general $LC(S) \le F_n^*(r_1,r_2,...,r_n)$, F_n^* is the integer function corresponding to F_n s.t. $F_n^*:Z^+ \rightarrow Z^+$. Since the 2^{nd} and 3^{rd} conditions are hold, then: $LC(S)=F_n^*(r_1,r_2,...,r_n)$

Linear complexity (LC) Criterion

Notice that LC(S) depends on LFSR and CF units. The basic condition to construct efficient KG is "Lengths of combined LFSR's must be long as possible". This condition will contributes to make S has maximum period. The other condition is "CF has high non-linear order", so if the five conditions are holding, this will make S has a high LC to pass the computer ability in exhaustive search or brute forces attack.

Now when applying the LC criterion on the studied cases we get:

1. **n-LKG**: S has $LC(S)=\sum_{i=1}^{n} r_i$.
2. **n-PKG**: S has $LC(S)=\prod_{i=1}^{n} r_i$.
3. **3-BKG**: S has $LC(S)=r_1 \cdot r_2 + r_1 \cdot r_3 + r_2 \cdot r_3$.

Example: Table describes LC for different examples of the three study cases

n	r_i	LC(S)		
		n-LKG	n-PKG	3-BKG
3	2,3,5	10	30	31
3	4,5,7	16	140	83
4	2,3,5,7	17	210	-----

Correlation Immunity (CI) Criterion

Definition: Let $X_1,X_2,...,X_n$ be independent binary variables, each taking on the values 0 or 1 with probability 1/2. A Boolean function $f(x_1,x_2,...,x_n)$ is m^{th}-order correlation immune if for each subset of m random variables $X_{i_1}, X_{i_2}, ..., X_{i_m}$ with $1 \le i_1 < i_2 < \ldots < i_m \le n$, the random variable $Z=f(X_1,X_2,...,X_n)$ is statistically independent of the random vector $(X_{i_1}, X_{i_2}, ..., X_{i_m})$.

Remark: If a Boolean function $f(x_1,x_2,...,x_n)$ is m^{th}-order CI, where $1 \le m < n$, then the nonlinear order of f is at most $n-m$. Moreover, if f is balanced then the nonlinear order of f is at most $n-m-1$ for $1 \le m \le n-2$.

The tradeoff between high LC and high CI can be avoided by permitting memory in the nonlinear combination function f.

For combination generators, the correlation attack can be prevented by using a CF f whose output is not correlated to any of its inputs. Such functions are called (n-1)-order correlation-immune.

For the 3-LKG, the CI(S)=n-1=2, but for the 3-PKG, the CI(S)=0, since the non-linear order of product system is 3.

Correlation Immunity (CI) Criterion

The CI order can be calculated from logical truth table for CF depending on calculating correlation probability, notice that CI depends on CF unit only and there is little effect of LFSR unit. Therefore, the condition to obtain efficient KG's **"Choosing CF with maximum order correlation immune"**, this condition is not essential since the correlation (if it exist) can prevented by using some ways. Moreover, Staffelbach, mentioned that to prevent correlation attack (the other condition to obtain efficient KG's) **"Using long LFSR's with maximum tapping number of connection polynomial"**.

System	Input			Output		
	x_1	x_2	x_3	F_L	F_P	F_B
	0	0	0	0	0	0
	0	0	1	1	0	0
	0	1	0	1	0	0
	0	1	1	0	0	1
	1	0	0	1	0	0
	1	0	1	0	0	1
	1	1	0	0	0	1
	1	1	1	1	1	1
Linear	0.5	0.5	0.5	**Correlation Probability**		
Product	0.625	0.625	0.625			
Brüer	0.75	0.75	0.75