

CRYPTOGRAPHIC PROPERTIES OF GROTH SEQUENCES

E. Dawson*, J. Asenstorfer**, P. Gray **

* School of Mathematics,
Queensland University of Technology.

**School of Electronic Engineering
South Australia Institute of Technology.

ABSTRACT

In a stream cipher the terms of an infinite binary sequence are added modulo two to binary plaintext to form ciphertext. One method proposed by Groth for forming this sequence is to sum second order products of the stages of a maximal length sequence where the products are defined by a Langford arrangement. It was claimed that a sequence could be produced in this fashion with maximum linear complexity provided sufficient layers are used. In this paper these sequences are analysed by using recent results on stream ciphers.

1. INTRODUCTION

One method of forming a binary sequence for use in a stream cipher is to apply a nonlinear function to the stages of a linear feedback shift register (LFSR) whose characteristic polynomial is primitive. In [1] Groth described a method of forming such a sequence by summing second order products at several layers. At each layer these products are defined by a Langford arrangement. For the remainder of this paper such a sequence will be referred to as a Groth sequence. It was claimed in [1] that for such a sequence the linear complexity can be expressed by a formula.

In Section 2 a description of the basic properties of stream ciphers will be given including the formation of a binary sequence by applying a nonlinear function to the stages of a maximal length LFSR. It will be shown how such a function can be defined by using the algebraic normal form from [2]. In Section 3 a definition of a Langford arrangement and a Groth sequence will be given from [1]. In Section 4 the cryptographic properties of Groth sequences will be examined. In Section 5 these properties will be given in relation to several examples of Groth sequences. Several of the results from Sections 4 and 5 have been given previously in [3] by the authors.

2. STREAM CIPHERS

An LFSR is defined by a recurrence relation

$$s_{j+L} = \sum_{i=0}^{L-1} c_i s_{i+j} \quad \text{for } j \geq 0 \quad (1)$$

where the c_i are binary constants such that $c_0 = 1$. Associated with such a recurrence relation is a polynomial

$$f(x) = c_0 + c_1x + \dots + c_{L-1}x^{L-1} + x^L \quad (2)$$

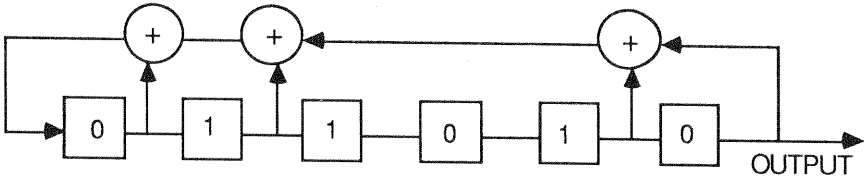


Figure 1. A block diagram of a LFSR.

which is defined to be the characteristic polynomial of the LFSR. The coefficients of $f(x)$ are called the feedback coefficients of the LFSR. A binary sequence (s_n) whose n th term is s_n , which satisfies equation 1 is defined by its initial state vector $(s_0, s_1, \dots, s_{L-1})$. For each (s_n) which satisfies equation 1 there exists a smallest positive number p such that $s_{n+p} = s_n$ for all n . The number p is said to be the period of the sequence. The LFSR with characteristic polynomial $1 + x + x^4 + x^5 + x^6$ is shown in Figure 1. If the initial state vector of the LFSR is $(0, 1, 0, 1, 1, 0)$ the resulting sequence has period 63 given by

$$01011001010100100111100000110111001100011101011111011010001000 \quad (3)$$

A stream cipher is the process of encryption where the terms of an infinite binary sequence are added modulo two to binary plaintext to form ciphertext. In this paper all such sequences will be periodic. In order to be secure from cryptanalysis by an attacker the sequence should satisfy three properties as described in [4].

Property 1. The period of the sequence should be large.

Property 2. The linear complexity of the sequence should be large.

Property 3. The sequence should have noise-like characteristics.

Every periodic sequence can be produced by an LFSR [2]. In general there are many LFSR's which can produce a given sequence. The degree of the shortest LFSR which will produce a sequence is defined to be the linear complexity of the sequence. For example the sequence given in (3) has a linear complexity of six since the LFSR in Figure 1 is the shortest LFSR which will produce this sequence. If the linear complexity of a sequence is L , then the feedback coefficients for the sequence can be found by applying the Berlekamp-Massey algorithm [5] provided $2L$ consecutive terms of the sequence are

known. Given the feedback coefficients and L consecutive terms the entire sequence can be derived by substitution into (1).

In [6] there are three different properties to measure the noiselike characteristics of a binary sequence. A sequence which satisfies these three properties is said to be G-random. In this paper only the first of these properties will be used to measure the randomness of a binary sequence. This property states that in a period of the sequence approximately one half of the terms should be a one.

A polynomial $f(x)$ over $GF(2)$ of degree L is said to be a primitive polynomial if $f(x)$ divides $x^{2^L-1} + 1$ but does not divide $x^k + 1$ for $k < 2^L-1$. By using an LFSR whose characteristic polynomial is primitive of degree L one can create a binary sequence of period 2^L-1 . Such a sequence is called a maximal length or an m-sequence. For example, the polynomial $1 + x + x^2 + x^5 + x^6$ is primitive, so that the LFSR in Fig. 1 generates an m-sequence of period 63.

An m-sequence is known to be G-random [6]. However, one would not use an m-sequence by itself in a stream cipher to encrypt binary plaintext since the linear complexity is small in comparison to the period length.

One method of forming a sequence (z_n) of large linear complexity is to use a nonlinear Boolean function f to combine the stages of a maximal length LFSR whose length is L . Let z be the binary vector of length 2^L-1 corresponding to the first 2^L-1 terms of (z_n) . Let s_1 be the binary vector of length 2^L-1 corresponding to the first 2^L-1 terms of the m-sequence used to define (z_n) . Let s_2, \dots, s_L be cyclic shifts of s_1 where s_2 corresponds to a shift of one place to the left and so forth. In [2] it was explained why the set, T , of 2^L-1 binary vectors of length 2^L-1 defined by

$$T = \{s_1, \dots, s_L, s_1 s_2, \dots, s_1 s_2 s_3, \dots, s_1 s_2 \dots s_L\} \quad (4)$$

is a basis of the vector space of all binary 2^L-1 tuples. Since T is a basis, we can write z as

$$z = a_1 s_1 + \dots + a_L s_L + a_{12} s_1 s_2 + \dots + a_{12\dots L} s_1 s_2 \dots s_L. \quad (5)$$

The expression for z in (5) is called the algebraic normal form of (z_n) . The algebraic order of (z_n) is the largest value of j such that there exists a non zero coefficient $a_{i_1 i_2 \dots i_j}$. For example the sequence defined by $z = s_1 s_2 + s_1 s_2 s_4$ has an algebraic order of three.

There are two theorems from [2] to define the cryptographic properties of the sequence (z_n) as given by (5).

Theorem 1. The period p of (z_n) is 2^L-1 or a divisor of 2^L-1 .

Theorem 2. The linear complexity of (z_n) is less than or equal to $\sum_{i=1}^r \binom{L}{i}$

where r is the algebraic order of the Boolean function defining the sequence.

A sequence (z_n) defined by (5) with algebraic order r is said to be degenerate if its linear complexity is less than $\sum_{i=1}^r \binom{L}{i}$.

3. GROTH SEQUENCES

The Langford problem [7] is to arrange numbers $1, 1, 2, 2, 3, 3, \dots, g, g$ in a sequence in such a way that for $h = 1, 2, 3, \dots, g$ the two h 's are separated by exactly h places; for example 41312432 ($g = 4$). We will call such a sequence a Langford arrangement. The number of Langford arrangements has an exponential-like increase with the size of g . The number of Langford arrangements has been found for all values of g less than or equal to 12. This is listed in Table 1, from [1] where arrangements that arise from reversals are omitted.

Table 1

Register length L	Multipliers	Langford Arrangements
6	3	1
8	4	1
14	7	26
16	8	150
22	11	17792
24	12	108144

A Groth sequence is produced by summing second-order products of the L stages of a maximal-length LFSR. If $L = 2g$ then these products are defined by a Langford arrangement on the numbers $1, 1, 2, 2, 3, 3, \dots, g, g$. The numbers in the Langford arrangement are assigned sequentially to the register stages and the two inputs of a multiplier are connected to the two stages with the same number. For example let $L = 6$. A Langford arrangement with three multipliers is 231213. Figure 2 indicates the Groth sequence which results if this arrangement is applied to the stages of the LFSR from Fig. 1.

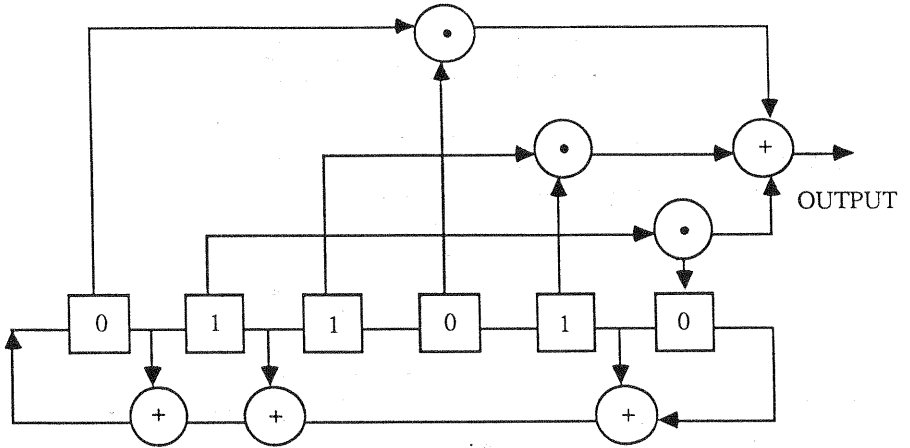
The algebraic normal form for this sequence as defined by (5) is

$$s_1 s_5 + s_2 s_4 + s_3 s_6.$$

A more complicated Groth sequence can be formed by using multilayers where second order products as defined by a Langford arrangement are summed at each layer as indicated in Figure 3.

The Langford type arrangement can be found for cases which are similar to the true Langford problem. For example in the case of $L = 12$ and $g = 6$ the $0, 0$ pair can be included by excluding the $1, 1$ pair. A Langford type arrangement that results is 005623425364. In the case where L is odd we can also create a Langford type arrangement. For example for $L = 7$ a Langford type arrangement would be a231213. In this case the non-linear function to create the Groth sequence would contain one first order term in the summation (as indicated by the letter a above) along with three second

order products. Hence, we can use such arrangements to build Groth sequences in the case where L is such that no true Langford arrangement exists.



101100111100101111000000110111110001000110100100001010010000001

Figure 2.

Groth sequence which results if Langford arrangement is applied to the stages of the LFSR from Fig. 1.

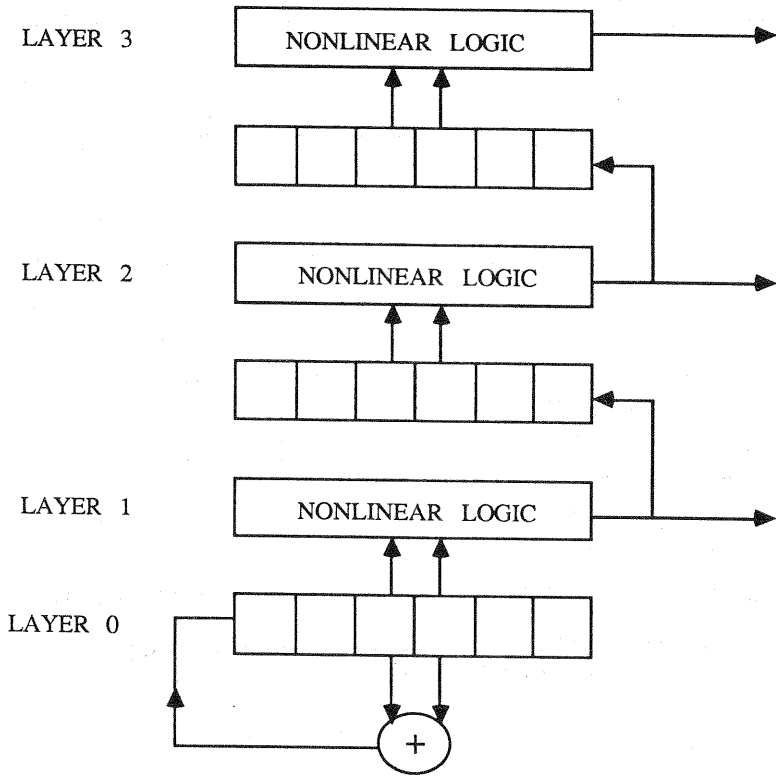


Figure 3.
Three layer multiplier arrangement of a Groth sequence

4. PROPERTIES OF GROTH SEQUENCES

4.1 Period

By using Theorem 1 a Groth sequence formed by an LFSR of length L has a period of length $2^L - 1$ or a divisor of $2^L - 1$. In order to guarantee that a Groth sequence has a period of length $2^L - 1$ one needs to use a Mersenne prime. If $2^L - 1$ is a prime number then $2^L - 1$ is classified as a Mersenne prime. The values of L less than 100 which produce Mersenne primes are 2, 3, 5, 7, 13, 19, 31, 61, 87 [8].

4.2 Linear Complexity

Let a Groth sequence be formed by an LFSR of length L . In [1] it was claimed that at each layer the linear complexity of the resulting Groth sequence can be expressed by a formula.

According to this claim the linear complexity at layer q will be

$$\sum_{i=1}^{2^q} \binom{L}{i}$$

provided $2^q < L$. If $L \leq 2^q < 2L$ the linear complexity claimed in [1] is the maximum value possible which is

$$\sum_{i=1}^L \binom{L}{i} = 2^L - 1.$$

For $L = 6$ the value of q to satisfy the above inequality is $q = 3$. Figure 3 displays the three layer multiplier arrangement of the Groth sequence which results from applying this configuration to the LFSR in Fig. 1. Using Groth's claim the linear complexity at each layer in Fig. 3 is listed in Table 2.

Table 2

Layer	Linear Complexity
1	21
2	56
3	63

In Section 5 the actual value of the linear complexity at each layer will be computed.

In order to state that the linear complexity of a Groth sequence is

$$\sum_{i=1}^r \binom{L}{i}$$

the algebraic order of the Boolean function as defined by (5) must be at least r . As will be shown in Section 5 both the algebraic order and the linear complexity can vary for different Groth sequences at the same level with the same length of LFSR depending on the choice of the primitive polynomial defining the tap settings and the Langford arrangement used at each level.

4.3 Noiselike Characteristics

The number of ones for a Groth sequence can be determined by applying some results from [9] provided the algebraic normal form is known. However as discussed in [3] the algebraic normal form for a Groth sequence at the layers above the first layer (where the idea of layers is as shown in Figure 3) is difficult to derive. Theorems 3 and 4 give the number of ones in a Groth sequence at the first layer for an LFSR of length L depending on whether or not L is an even or odd number respectively.

Theorem 3 [3]. The number of ones in each 2^{L-1} consecutive terms is $2^{r-1}(2^r-1)$ when $L = 2r$.

Theorem 4 [10]. If L is odd the number ones in each 2^{L-1} consecutive terms is 2^{L-1} or $2^{L-1}-1$.

As will be shown in Section 5 the number of ones in 2^{L-1} consecutive terms can vary for Groth sequences defined at above the first layer.

4.4 Crosscorrelation Function

In [11] it is demonstrated how a sequence defined in the form of (5) can be attacked by using the crosscorrelation function between the m-sequence and the produced sequence. In this attack the cryptanalyst is assumed to know the characteristic polynomial of the defining m-sequence and a small amount of ciphertext. Both the initial state vector used to form the m-sequence and the Boolean function are unknown to the cryptanalyst. The attack consists of deriving an equivalent system.

For the attack mentioned above to be successful in a reasonable amount of time only a small number of stages of the LFSR defining the m-sequence must be used in the formation of the Boolean function. A Groth sequence as defined in Section 3 is secure from this crosscorrelation attack since all the stages of the defining m-sequence are used in the formation of the sequence.

4.5 Key Size

If an LFSR of length L is used to form a Groth sequence there are three different sources of key.

Source 1. Initial State Vectors.

There are 2^{L-1} possible nonzero initial state vectors for an LFSR of length L .

Source 2. Tap Settings.

Let $\phi(n)$ denote the number of positive integers less than n which are relatively prime to n .

There are $\lambda(L) = \frac{\phi(2^L-1)}{L}$ from [4] primitive polynomials of degree L .

Table 3 indicates the exponential like increase in $\lambda(L)$.

Table 3.

L	$\lambda(L)$	L	$\lambda(L)$	L	$\lambda(L)$
1	1	9	48	17	7710
2	1	10	60	18	8064
3	2	11	176	19	27594
4	2	12	144	20	24000
5	6	13	630	21	84672
6	6	14	756	22	120032
7	18	15	1800	23	356960
8	16	16	2048	24	276480

Source 3. Langford Arrangements.

As shown by Table 1 there is an exponential like increase in possible Langford arrangements to use at each layer. In [1] there is an algorithm for generating Langford arrangements.

As an example of the key size let $L = 24$ and suppose that a Groth sequence is formed using five layers of products. For this sequence there are:

- approximately 1.6×10^7 initial state vectors,
- approximately 2.8×10^5 primitive polynomials,
- approximately 2.2×10^5 Langford arrangements.

Hence the total number of possible keys is:

$$(1.6 \times 10^7) (2.8 \times 10^5) (2.2 \times 10^5)^5 \approx 1.8 \times 10^{38}.$$

In comparison the DES algorithm has approximately 6.5×10^{16} keys.

5. EXAMPLES OF GROTH SEQUENCES

For an LFSR of length six there are six possible primitive polynomials. There are two Langford arrangements of length, six, 312132 and 231213. Let 0 and 1 stand for the Langford arrangements 312132 and 231213 respectively. For a given primitive polynomial one can generate

- (a) two Groth sequences at layer one by using Langford arrangements denoted by 0 or 1;
- (b) four Groth sequences at layer two by using Langford arrangements denoted by 00, 01, 10, 11;
- (c) eight Groth sequences at layer three by using Langford arrangements denoted by 000, 001, 010, 011, 100, 101, 110, 111.

All together one can generate 84 Groth sequences by using an LFSR of length six. Tables 4 to 6 contain a list of properties of these sequences together with the Langford arrangements that were used. The linear complexity was derived by applying the Berlekamp-Massey algorithm. As mentioned previously, according to Groth's claim the linear complexities of these sequences at layers 1, 2, 3 would be 21, 56, 63 respectively from Table 2. By inspection of Table 6 there is an increase in linear complexity at each layer although many of the sequences are degenerate. As shown by Table 6 the algebraic order varies for the Groth sequences at the third level varies between five and six.

Let LFSR1 have a primitive polynomial $f(x)$ of degree L as a characteristic polynomial. Let LFSR2 have $f^*(x)$ as characteristics polynomial where $f^*(x) = x^L f(1/x)$ is called the reciprocal polynomial of $f(x)$. It can be shown that $f^*(x)$ is also a primitive polynomial. Let (z_n) be a sequence produced by applying a non-linear filter function f to LFSR1. Let (y_n) be a sequence produced by applying the reciprocal function (which corresponds to symmetrically changing the function taps) to LFSR2. In [3] it is shown that (y_n) is a reversal sequence of (z_n) (if the first period vector of (z_n) is $[z_1, \dots, z_p]$ then the first period vector of (y_n) is $[z_p, \dots, z_1]$), and it is shown that z_n and y_n have the same linear complexities. For example the polynomial $1+x+x^6$ is the reciprocal polynomial of $1+x^5+x^6$. Hence the Groth sequences generated by $1+x+x^6$ are reversal sequences of the Groth sequences generated by $1+x^5+x^6$. In Tables 4 to 6 we omitted the properties of the reversal sequences generated by the reciprocal polynomials.

Table 4

Properties for Groth Sequences for different primitive polynomials and Langford arrangements (Layer 1)

Primitive polynomial	Langford arrangement	Linear complex	No. zeros	Algebraic order
$1+x^5+x^6$	0	21	35	2
	1	21	35	2
$1+x^2+x^3+x^5+x^6$	0	21	35	2
	1	21	35	2
$1+x^2+x^4+x^5+x^6$	0	21	35	2
	1	18	35	2

Table 5

Properties for Groth Sequences for different primitive polynomials and Langford arrangements (Layer 2)

Primitive polynomial	Langford arrangement	Linear complex	No. zeros	Algebraic order
$1+x^5+x^6$	00	56	39	4
	01	56	37	4
	10	53	37	4
	11	53	35	4
$1+x^2+x^3+x^5+x^6$	00	56	35	4
	01	56	41	4
	10	56	41	4
	11	56	41	4
$1+x^2+x^4+x^5+x^6$	00	56	39	4
	01	56	41	4
	10	54	41	4
	11	48	39	4

Table 6

Properties for Groth Sequences for different primitive polynomials and Langford arrangements (Layer 3)

Primitive polynomial	Langford arrangement	Linear complex	No. zeros	Algebraic order
$1+x^5+x^6$	000	60	45	5
	001	60	45	5
	010	56	51	5
	011	62	49	5
	100	61	40	6
	101	61	40	6
	110	62	33	5
	111	59	33	5
$1+x^2+x^3+x^5+x^6$	000	56	45	5
	001	62	45	5
	010	59	39	5
	011	62	41	5
	100	61	44	6
	101	61	44	6
	110	62	47	5
	111	62	45	5
$1+x^2+x^4+x^5+x^6$	000	57	42	6
	001	63	42	6
	010	62	43	5
	011	56	43	5
	100	62	49	5
	101	62	51	5
	110	60	46	6
	111	60	50	6

A Groth sequence at each layer for a shift register of length 7 to 13 was generated and analysed. A list of the properties of these sequences together with the Langford arrangements that were implemented are included in Table 7.

This table includes the actual linear complexity and the linear complexity which was claimed by Groth. By inspection the linear complexities are close to the value claimed although several of the sequences are degenerate. The period length for each sequence is included. As indicated by the number of zeros in relation to the period length the noiselike characteristics of each sequence are reasonable.

Table 7

Property of Groth Sequences with LFSR length
7 to 13 with various Langford arrangements

Primitive polynomial	Langford arrange-	Layer Complex	Actual Linear Complex	Claimed Linear	Period	No. zeros
x^7+x^6+1	312132a	1	28	28	127	63
	213213a	2	98	105	127	55
	312132a	3	126	127	127	63
$x^8+x^7+x^3+x^2+1$	41312432	1	36	36	255	135
	23421314	2	162	162	255	147
	41312432	3	255	255	255	160
x^9+x^5+1	41312432a	1	45	45	511	255
	23421314a	2	255	255	511	263
	41312432a	3	501	510	511	247
	23421314a	4	511	511	511	248
$x^{10}+x^7+1$	5004235243	1	55	55	1023	527
	3425324005	2	385	385	1023	543
	5004235243	3	1007	1012	1023	585
	3425324005	4	1018	1023	1023	616
$x^{11}+x^9+1$	5004235243a	1	66	66	2047	1023
	3425324005a	2	561	561	2047	1047
	5004235243a	3	1980	1980	2047	1053
	3425324005a	4	2047	2047	2047	1018
$x^{12}+x^9+x^8+x^5+1$	640053462352	1	78	78	4095	2079
	500463524326	2	793	793	4095	2103
	004635243265	3	3792	3796	4095	2105
	004562342536	4	4092	4095	4095	2156
$x^{13}+x^{12}+x^{10}+x^9+1$	005623425304a	1	91	91	8191	4095
	005623425364a	2	1092	1092	8191	4143
	005623425364a	3	7098	7098	8191	4125
	005623425364a	4	8190	8191	8191	4053

CONCLUSION

As was demonstrated in Section 5 by counter-example the claim from [1] for the linear complexity of a Groth sequence is false. However, these examples showed that after increasing the length of the defining LFSR the linear complexity was close to Groth's original claim. This supports the argument from [2] that if the algebraic order of the Boolean function is r for a sequence formed as in (5) from an LFSR of length L one can state with a probability close to one that the linear complexity is approximately

$$\sum_{i=1}^r \binom{L}{i}$$

provided L is sufficiently large. Furthermore a large value of L will provide a large key size for a Groth sequence as was demonstrated in Section 4 and the noiselike characteristics for such a sequence should be good in terms of the number of ones in a period as was shown by the examples in Section 5.

REFERENCES.

1. E.J. Groth, 'Generation of binary sequences with controllable complexity', Trans. IEEE, Vol. IT-17, May 1971, pp. 288-296.
2. R.A. Rueppel, 'Analysis and Design of Stream Ciphers', Springer-Verlag, Berlin, 1986.
3. J. Asenstorfer, E. Dawson, P. Gray, 'Analysis of Groth Sequences', Journal of Electrical and Electronics Engineering, Australia, Vol. 8, No. 4, Dec. 1988, pp. 211-221.
4. H. Beker and F. Piper, 'Cipher Systems: The Protection of Communications', Wiley Interscience, NY, 1982.
5. J.L. Massey, 'Shift register sequences and BCH Decoding', Trans. IEEE, Vol. IT-15, Jan. 1969, pp. 122-127.
6. S.W. Golomb, 'Shift Register Sequences', Aegean Park Press, Laguna Hills, California, 1982.
7. C.D. Langford, 'Problem', Math. Gaz., Vol. 42, Oct. 1958, p. 228.
8. D.E. Knuth, 'The Art of Computer Programming, Vol.2: Seminumerical Algorithms', Addison-Wesley, Reading, MA, 1969.
9. N. Kalouptsidis and M. Manolarakis, 'Sequences of Linear Feedback Shift Registers with Nonlinear-Feedforward Logic', Proc. IEE, Part E, Computers and Digital Techniques, Vol. 130, No. 5, 1983, pp. 174-176.
10. M.K. Simon, J.K. Omura, R.A. Scholtz, B.K. Levitt, 'Spread Spectrum Communications Vol.1', Computer Science Press, Maryland, 1985.
11. T. Siegenthaler, 'Cryptanalysis Representation of Nonlinearly Filtered ML-Sequences', Eurocrypt 85, Springer Verlag, 1986, pp. 103 - 110.

