Large Primes in Stream Cipher Cryptography

Kencheng Zeng[1], C.H. Yang[2], and T.R.N. Rao[2]

[1]Graduate School of USTC
Academia Sinica
P.O. Box 3908
Beijing
People's Republic of China

[2]The Center for Advanced Computer Studies
University of Southwestern Louisiana
P.O. Box 44330
Lafayette, LA 70504-4330
U.S.A.

Abstract. The present research is motivated by the observation that if the period T of a certain binary sequence is a prime, then its linear complexity will be bounded from below by the order of 2 modulo T, i.e., $LC \geqslant Ord_T(2)$. A class of generators with state periods $T(q, n) = q \cdot 2^n - 1$ are constructed for $q = 3, 5, 7, 9$ and arbitrary n on the basis of a pair of m-sequence generators with the same number of stages, each controlling the clock of the other (bilateral stop-and-go clock control). A new test is derived to find the primes among the numbers $T(q, n)$ with the cases $3 \mid q$ and $3 \nmid q$ treated in a unified manner. The orders of 2 modulo some of the primes $T(q, n)$ are given and some additional cryptographic and implementational remarks are made.

I. Introduction

In designing a qualified keystream generator, one of the first concerns of the cryptographer is to guarantee a large enough key-independent lower bound to the linear complexity of the output sequences. More desirably, if such a lower bound can be made to be of an order of magnitude approximately equal to that of the period.

* This research is supported by Board of Regents of Louisiana Grant #86-USL(2)-127-03

One of the traditional ways for attaining this objective [1] is to start with an n-stage LFSR which generates the driving sequence

$$\mathbf{a} = \{ a(0), a(1), \cdots, a(i), \cdots \},$$

and apply to it a nonlinear feedforward transformation of the form

$$b(i) = f_K(a(i), a(i-1), ..., a(i-r+1))$$

to produce the nonlinear feedforward sequence

$$\mathbf{b} = \{ b(0), b(1), \cdots, b(i), \cdots \},$$

where the transfer function $f_K(x_0, x_1, ..., x_{r-1})$ is selected by the key K from a certain family $\{f_K\}$ of algebraic polynomials, which are linear in each one of the r indeterminates separately. It can be shown [2], that the linear complexity LC of the feedforward sequence **b** is bounded from above by the inequality

$$LC \leqslant \sum_{k=1}^{d} C_k^d$$

where $d \leqslant r$ is the total degree of the transfer polynomial. This key-dependent upper bound can be attained only for carefully chosen transfer functions f_K and, according to Siegenthaler's theory [3], any attempt aimed at increasing d will lead to a decrease in the degree of correlation-immunity of f_K.

For de Bruijn sequences of period 2^n, we have LC $> 2^{n-1}$, but such sequences either are hard to implement technically or suffer from severe auto-correlation weakness [4].

In a recent work, Gollmann and Chambers [5, 6] proposed another interesting approach by considering a cascade of n clock controlled LFSRs of the same length $p \geqslant 3$, which is chosen to be a prime. The output sequence has been shown to have period p^n and a linear complexity of approximately the same order of magnitude. The drawback is that approximately $p \log T / \log p$ storage elements will be needed in order to produce a sequence of period T, in addition to the fact that several clock cycles will be needed for the generation of one single pseudorandom bit.

There is yet another simple approach which, to our knowledge, has been exploited nowhere for the purpose of achieving a guaranteed lower bound to the linear complexity. Namely, we have the following

Proposition. If a binary sequence **b** has an odd prime period T, then its linear complexity will be bounded from below by the order of the number 2 modulo T, i.e.,

$$LC(\mathbf{b}) \geqslant Ord_T(2) . \qquad (1)$$

Proof. Since T is odd, the minimum polynomial $\mu(x)$ of $\mathbf{b}$ must have an irreducible factor, say $p(x)$, of degree $d > 1$, and any root α of $p(x)$, as an element in the multiplicative group of its splitting field E, has order T. Thus, if we write $d^* \triangleq Ord_T(2)$, then

$$\alpha\ ,\ \alpha^2\ ,\ \cdots\ ,\ \alpha^{2^{d^*}-1}$$

will be distinct elements in E with $\alpha^{2^{d^*}} = \alpha$. This means

$$p^*(x) \triangleq \prod_{i=0}^{d^*-1} (x - \alpha^{2^i})$$

will be an irreducible polynomial in $F_2[x]$ with a root α in common with $p(x)$. So we must have $p(x) = p^*(x)$ and, in particular, $d = d^*$. consequently,

$$LC(\mathbf{b}) = deg\ (\mu(x)) \geqslant deg\ (p(x)) = d^* = Ord_T(2)\ .$$

A sequence with a prime period has another desirable property that one can subject it to various further cryptographic transformations without influencing the established lower bound to its linear complexity, provided the transformations do not render it into a constant sequence.

However, if $T = 2^n - 1$ happens to be one of the known Mersenne primes, then we shall have $Ord_T(2) = n$. So T should be chosen to be a prime not belonging to the progression $2^n - 1\ ,\ n \geqslant 1$. An idea which naturally arises in this connection is to consider generators with state periods which are primes of the form

$$T(q, n) = q \cdot 2^n - 1\ ,\ q > 1\ ,\ q \equiv 1\ mod\ 2\ .$$

In the sequel, we shall construct generators with state periods $T(q, n)$ for $q = 3, 5, 7, 9$ and arbitrary n. A new algorithm will be designed to find the primes among these numbers, and the orders of 2 modulo the primes thus found will be computed to show that this new approach will be successful.

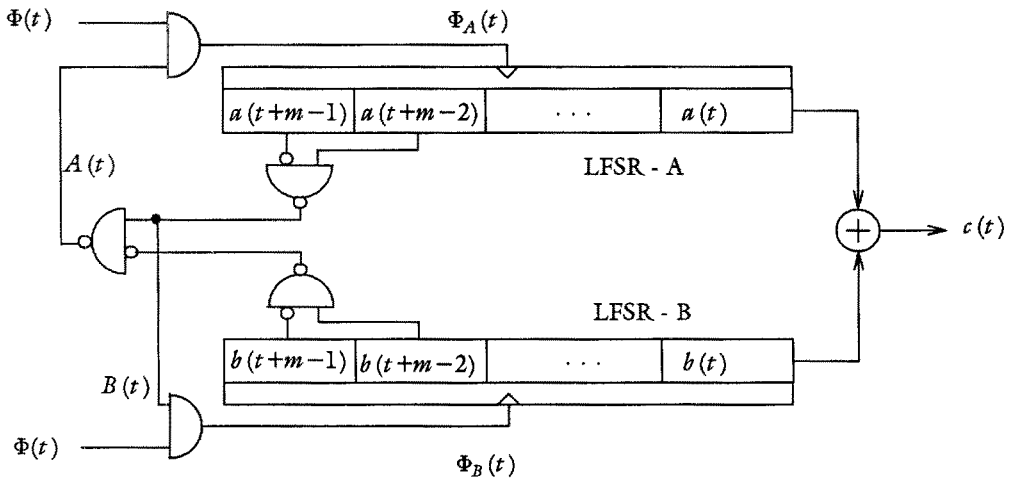
II. Generators with State Periods $T(q, n)$

We shall construct random-bit generators which will describe cycles of length $T(q, n)$ for $q = 3, 5, 7, 9$ and arbitrary n, in their corresponding spaces of admissible states. But for simplicity of exposition, we shall discuss in details only the case $q = 5$ and shall make an analysis of the state diagram of the generator thus constructed. The cases $q = 3, 7, 9$ can be treated with slight modifications.

In the scheme showed below, we have a pair of m-stage maximal length linear feedback shift registers LFSR-A and LFSR-B. Each of these LFSRs controls the clock pulses to the other in the following way:

(i) If $(a(t+m-1), a(t+m-2)) = (0,1)$, then the clock pulse to LFSR-B is to be blocked;

(ii) If $(b(t+m-1), b(t+m-2)) = (0,1)$, but $(a(t+m-1), a(t+m-2)) \neq (0,1)$, then the clock pulse to LFSR-A is to be blocked.



Thus, if we denote the t-th pulse from the clock by $\Phi(t)$, then the $(t+1)$-th pulses to LFSR-A and LFSR-B will be

$$\Phi_A(t+1) = A(t)\cdot\Phi(t+1)\,, \tag{2}$$

and

$$\Phi_B(t+1) = B(t)\cdot\Phi(t+1)\,, \tag{3}$$

where

$$B(t) = \overline{\bar{a}(t+m-1)a(t+m-2)} \tag{4}$$

and

$$A(t) = \overline{\bar{a}(t+m-1)a(t+m-2)\,\bar{b}(t+m-1)b(t+m-2)}\,. \tag{5}$$

The output signal $c(t)$ of the generator at the moment t is given by

$$c(t) = a(t) + b(t)\,. \tag{6}$$

The phase space of the generator consists of vector-pairs of the form $\mathbf{s} = (\mathbf{s_A}, \mathbf{s_B})$, where

$$\mathbf{s_A} \triangleq (a(t+m-1), a(t+m-2), \cdots, a(t))$$

$$\mathbf{s_B} \triangleq (b(t+m-1)\,,\, b(t+m-2)\,,\, \cdots\,,\, b(t))$$

are non-zero vectors of $V_m(F_2)$, so that the number of admissible states of the generator is $(2^m - 1)^2$. The state diagram, however, will compose of branched cycles, and we need the following minimum of working terminology in order to describe its structure.

Definition. We say the admissible state $\mathbf{s}$ *precedes* the state $\mathbf{s}^*$ or $\mathbf{s}^*$ *succeeds* $\mathbf{s}$, if our generator can go over from $\mathbf{s}$ to $\mathbf{s}^*$ in one step of work. We say a state $\mathbf{s}$ is *inaccessible* to the generator if it has no predecessor.

Besides, for any non-zero vector $\mathbf{s} \in V_m(F_2)$, we denote by $L_A\mathbf{s}$, or $L_B\mathbf{s}$, the vector which precedes $\mathbf{s}$ when the shift register LFSR-A, or LFSR-B, with undisturbed clock pulses is set to work. Similarly, we denote by $R_A\mathbf{s}$, $R_B\mathbf{s}$, the successors of $\mathbf{s}$ under the work of LFSR-A, LFSR-B, with undisturbed clock pulses.

Theorem 1. (i) There are 2^{2m-4} inaccessible states in the state diagram of the generator, and an admissible state $\mathbf{s} = (\mathbf{s_A}\,,\, \mathbf{s_B})$ is inaccessible to the generator if and only if

$$(a(t+m-1),\, a(t+m-2)) = (0,\, 1)\,, \quad (b(t+m-2),\, b(t+m-3)) = (0,\, 1) \tag{8}$$

(ii) Every branch to a cycle starts with an inaccessible state and has length 1, and at any cycle state there is at most one entering branch;

(iii) The accessible states in the state diagram fall in

$$T(3,\, m-2) = 3\cdot 2^{m-2} - 1 \tag{9}$$

cycles, each of length

$$T(5,\, m-2) = 5\cdot 2^{m-2} - 1\,. \tag{10}$$

Proof. (i) Assume (8) satisfied. If the state $\mathbf{s}$ has a predecessor, then it must be one of the states:

$$(L_A\mathbf{s_A}\,,\, \mathbf{s_B})\,,\ (\mathbf{s_A}\,,\, L_B\mathbf{s_B})\,,\ (L_A\mathbf{s_A}\,,\, L_B\mathbf{s_B})\,.$$

But the corresponding successors of these states under the work of our generator will be respectively

$$(\mathbf{s_A}\,,\, R_B\mathbf{s_B})\,,\ (R_A\mathbf{s_A}\,,\, L_B\mathbf{s_B})\,,\ (L_A\mathbf{s_A}\,,\, \mathbf{s_B})\,,$$

and none of them coincides with $\mathbf{s}$.

Conversely, if at least one of the two conditions in (8) is not satisfied, then similar arguments can be used to find a predecessor for $\mathbf{s}$.

(ii) To see this, we need only observe that no state $\mathbf{s} = (\mathbf{s_A}\,,\, \mathbf{s_B})$ can have more than two predecessors and it will have two if and only if

$$(a(t+m-2), a(t+m-3)) = (0, 1)\ ,\quad (b(t+m-2), b(t+m-3)) = (0, 1) \quad (11)$$

where the predecessors are

$$(L_A \mathbf{s_A}\ ,\ \mathbf{s_B})\ ,\quad (\mathbf{s_A}\ ,\ L_B \mathbf{s_B})\ ,$$

with the first one being inaccessible.

(iii) To see this, we observe that the clock-controlling sequences $\{A(t)\}$ and $\{B(t)\}$ have the following simple properties:

(a) If one of the two shift-registers, say LFSR-A, being in the state $\mathbf{s_A}$ at a certain moment t, returns to that state for the first time at another mement t^*, then the number of 0's in the segment

$$B(t), B(t+1), ..., B(t^*)$$

is 2^{m-2};

(b) In both sequences $A(t)$ and $B(t)$ all zero ranges are of length 1.

Now suppose the generator starts at the moment $t = 1$ from the cycle state $\mathbf{s} = (\mathbf{s_A}, \mathbf{s_B})$ and LFSR-A returns to $\mathbf{s_A}$ for the first time at moment t_1, while LFSR-B returns to $\mathbf{s_B}$ for the first time at moment t_2, $t_1 \leqslant t_2$. Then

$$t_1 = 2^m + x - 1\ ,$$

where x is the number of 0's in the segment

$$A(1), A(2), \cdots, A(t_1)\ ,$$

and LFSR-B will sweep over

$$t_1 - 2^{m-2} = 3 \cdot 2^{m-2} + x - 1$$

distinct m-vectors in the time interval $[1, t_1]$. But this will mean that in the interval $[t_1 + 1\ ,\ t_2]$ the shift-register LFSR-B has to sweep over the remaining $2^{m-2} - x$ non-zero m-vectors and, at the same time, give rise to the same number of 0's in the segment

$$A(t_1+1), A(t_1+2), ..., A(t_2)\ .$$

So we see from the property (b), that we must have $x = 2^{m-2}$ and hence

$$t_1 = t_2 = 2^m + 2^{m-2} - 1\ .$$

The number of distinct cycles in the state diagram follows immediately from the identity

$$(2^m - 1)^2 - 2^{2m-4} = (3 \cdot 2^{m-2} - 1)(5 \cdot 2^{m-2} - 1)\ .$$

III. Testing the primality of $T(q, n)$

Necessary and sufficient conditions for numbers of the form $T(q, n)$ with arbitrary q to be prime have been discussed in [7, 8, 9, 10], the

computations needed are not expensive, but the cases $3 \mid q$ and $3 \nmid q$ have to be treated separately and the deduction of the conditions is not simple enough as to be put down in a single page. Since in our case q has only one prime factor, we would like to use a new test which, though computationally a little more expensive, treats the two cases in a unified way and can be derived immediately from the elegant elementary proposition of Lenstra [11] quoted below.

Proposition (Lenstra 1982). Let N and s be positive integers, and let A be a commutative ring with 1 containing $Z/(N)$ as a subring (with the same 1). Suppose that there exists $\alpha \in A$ satisfying the following conditions:

(1) $\alpha^s = 1$,

(2) $\alpha^{s/p} - 1 \in A^*$ (the group of units of A) for every prime p dividing s,

(3) the polynomial

$$f(x) = \prod_{i=0}^{t-1} (x - \alpha^{N^i})$$

has coefficients in $Z/(N)$ for some positive integer t.

Then every divisor r of N is congruent to a power of N modulo s.

Theorem 2. Assume $n \geqslant 2$, $q = \pi^a$ is a prime power, and let b be any integer. Define three sequences of integers by the recursive relations:

$$W_{i+1} = bW_i + W_{i-1}, \quad W_0 = 2, \quad W_1 = b, \quad 1 \leqslant i \leqslant q-1; \tag{12}$$

$$U_{i+1} = U_i^2 - 2, \quad U_1 = W_q^2 + 2, \quad 1 \leqslant i \leqslant n-2; \tag{13}$$

$$V_{i+1} = V_i^2 - 2, \quad V_1 = W_{\pi^{a-1}}^2 + 2, \quad 1 \leqslant i \leqslant n. \tag{14}$$

(i) If $T(q, n)$ is a prime, of which b^2+4 is a quadratic nonresidue (QNR), then

$$U_{n-1} \equiv 0 \ mod \ T(q, n); \tag{15}$$

(ii) If $U_{n-1} \equiv 0 \ mod \ T(q, n)$ and

$$(V_{n+1} - 2, T(q, n)) = 1, \tag{16}$$

then $T(q, n)$ is a prime.

Proof. Write

$$f(x) = x^2 - b\,x - 1$$

and let α and $\beta = b - \alpha$ be its two roots in the ring $A = Z/(f(x))$.

Evidently, we have

$$W_i = \alpha^i + \beta^i ,$$

$$U_i = \alpha^{q\,2^i} + \beta^{q\,2^i} ,$$

$$V_i = \alpha^{\pi^{a-1}2^i} + \beta^{\pi^{a-1}2^i}$$

(i) In this case, we can replace Z by the prime field $F = Z/(T(q, n))$ and A by the extension of F obtained by adjoining to it the root α of the irreducible quadratic $f(x)$. As is well-known, we must have $\beta = \alpha^{T(q,n)}$ which, together with $\alpha\beta = -1$, means $\alpha^{q\,2^n} = -1$, and hence

$$U_{n-1} = \alpha^{q\,2^{n-1}} + \beta^{q\,2^{n-1}} = \beta^{q\,2^{n-1}}(\alpha^{q\,2^n} + 1) = 0$$

(ii) From $U_{n-1} = 0$ we get $\alpha^{q\,2^n} = -1$, and hence

$$\beta = -\alpha^{-1} = \alpha^{q\,2^n - 1} = \alpha^{T(q,n)} .$$

Thus if we write $N \triangleq T(q, n)$ and $s \triangleq q\ 2^{n+1}$, then we shall have

(1) $\alpha^s = \alpha^{q\ 2^{n+1}} = 1$;

(2) For the two prime factors 2 and π of s, we have

(2.1). $\alpha^{s/2} - 1 = \alpha^{q\,2^n} - 1 = -2 \in A^*$,

(2.2). $\alpha^{s/\pi} - 1 = \alpha^{\pi^{a-1}2^{n+1}} - 1 = \ \in A^*$, for we have

$$(\alpha^{\pi^{a-1}2^{n+1}} - 1)(\beta^{\pi^{a-1}2^{n+1}} - 1) = 2 - (\alpha^{\pi^{a-1}2^{n+1}} + \beta^{\pi^{a-1}2^{n+1}}) = 2 - V_{n+1} \in A^* .$$

(3) Moreover, we see

$$f(x) = (x - \alpha)(x - \beta) = (x - \alpha)(x - \alpha^N)$$

has coefficients belonging to $Z/(N)$.

Therefore, we see from the proposition of Lenstra, every divisor r of N will be congruent to 1 or N modulo s. But

$$s = q\,2^{n+1} > q\,2^n - 1 = N ,$$

so r must be either 1 or N, which proves the primality of N.

We use (15) to discard the composites and use (16) to confirm the primality of those remain. But (ii) is only a sufficient condition, so in some cases several different b's have to be examined.

Algorithm. Consider a suitably long sequence of integers b_i, $i \geqslant 1$, such that $p_i = b_i^2 + 4$ are primes. The first ones of them are

$$5, 13, 29, 53, 173, 229, 293, 733 \cdots$$

Given the number $T(q, n)$, the primality test proceeds in the following way, with all computations carried out modulo $T(q, n)$:

Step 1. $i \leftarrow 0$.

Step 2. $i \leftarrow i+1$, check whether

$$\left(\frac{T(q, n) \bmod p_i}{p_i}\right) = -1 . \tag{17}$$

If yes, go to Step 3. Otherwise repeat Step 2.

Step 3. Check whether (15) is true. If yes, go to Step 4. Otherwise T is composite.

Step 4. If $V_{n+1} - 2 \equiv 0 \bmod T$, then return to Step 2.

Step 5. Check whether (16) is true. If yes, then $T(q, n)$ is prime. Otherwise T is composite.

Since every p_i is a prime of the form $4k+1$, by the Gaussian law of reciprocality, (17) means $p_i \in QNR \bmod T(q, n)$, as required in (i). The functions

$$\left(\frac{x}{p_i}\right), \quad 0 \leqslant x \leqslant p_i - 1 ,$$

are stored to facilitate the work of Step 2.

The following is a table of all primes of the form $T(q, n)$ with $q = 3, 5, 7, 9$ and $n \leqslant 1000$.

q	n										
3	1	2	3	4	6	7	11	18	34	38	43
	55	64	76	94	103	143	206	216	306	324	391
	458	470	827								
5	2	4	8	10	12	14	18	32	48	54	72
	148	184	248	270	274	420					
7	1	5	9	17	21	29	45	177			
9	1	3	7	13	15	21	43	63	99	109	159
	211	309	343	415	469	781	871	939			

For the cryptographically most interesting case where $q = 5$ the numbers $T(q, n)$ were found to be prime for the following values of $n \leqslant 4000$

1340, 1522, 1638, 2014, 2170, 2548, 2622, 2652, 2704.

IV. Computing $Ord_{T(q, n)}(2)$

We are interested only in the case where $T(q, n)$ is a prime. Moreover, when $n \geqslant 3$, the number 2 is a quadratic residue modulo $T(q, n)$, so in

computing the order of 2 modulo $T(q, n)$ we need only consider the complete factorization of

$$T(q, n-1) = \frac{T(q, n) - 1}{2} .$$

The computations are carried out only for $q = 5$ with $n \leqslant 150$, the case which seems to be cryptographically more interesting than the cases $q = 3, 7, 9$. Instead of $Ord_{T(q,n)}(2)$, the index of the subgroup generated by 2 in the multiplicative group of integers modulo the prime $T(q, n)$ is given in the following table, together with the smallest primitive root modulo that prime.

n	number of digits	index	smallest primitive root
4	2	2	2
8	4	2	3
10	4	2	3
12	5	2	3
14	5	2	3
18	7	6	3
32	11	2	3
48	16	2	3
54	17	2	6
72	23	2	6
148	46	78	7

The computation results meet the expectation stated in Section I satisfactorily.

V. Some Cryptographic Remarks

The point, however, is not in the linear complexity alone. It is a common practice to produce strong keystreams by combining up a certain number of comparatively simple source sequences, and the problem is that the complexities carried by the individual components are in many cases not well alloyed, so that an attack applied to one component may reveal the secrecy in the remaining parts. Take, for example, the so-called multiplexing scheme proposed in [12]. The whole system can be cracked either by the linear consistency test of [13], applied to the LFSR which provides the address numbers, or, as suggested in [14], by the correlational attack applied to the LFSR which provides the signals to be selected.

The idea of bilateral clock-control as describe in the present paper provides a good approach to intensifying the multiplexing scheme. If in our scheme, instead of (6) we produce the output signal $c(t)$ according to the multiplexing scheme with LFSR-A providing the address numbers and

LFSR-B providing the signals to be selected, then the attacks given in [13, 14] will be bound to fail for the two shift-registers are now made inseparable from each other, so that in attacking to one of them the cryptanalyst must take into consideration the other which controls the clock signals to it. The occurrence of a constant output sequence may be avoided by some threshold monitoring device, and so the lower bound to the linear complexity established before will hold true for the new scheme.

References

1. R.A. Rueppel, *Analysis and Design of Stream Ciphers*, Springer-Verlag, 1986.

2. Z.D. Dai et al., "Nonlinear Feedforward Sequences of m-sequences," *Proceedings of Beijing International Workshop of Information Theory*, 1988.

3. T. Siegenthaler, "Correlation-Immunity of Nonlinear Combining Functions for Cryptographic Applications," *IEEE Trans. on Info. Theory*, Vol. **IT-31**, Sep. 1984, pp. 776-780.

4. Z.D. Dai, "On the Construction and Cryptographic Applications of de Bruijn Sequences," submitted to *Journal of Cryptology*, 1989.

5. D. Gollmann and W. G. Chambers, "Stepping Clock Controlled Shift Registers," *EUROCRYPT 89*.

6. D. Gollmann and W.G. Chambers, "Clock-Controlled Shift Registers: A Review," *IEEE J. on Selected Areas in Comm.*, Vol. 7, 1989, pp. 525-533.

7. D.H. Lehmer, "An Extended Theory of Lucas' Functions," *Annals of Math.*, Vol. **31**, 1930, pp. 419-448.

8. H. Riesel, "A note on prime Numbers of the forms $N = (6a+1)2^{2n-1} - 1$ and $M = (6a+1)2^{2n} - 1$," *Ark. för Mat.*, 1955, pp. 245-253.

9. H. Riesel, "Lucasian Criteria For the Primality of $N = h \cdot 2^n - 1$," *Math. Comp.*, Vol. **23**, 1969, pp. 869-875.

10. H. Riesel, *Prime Numbers and Computer Methods for Factorization*, Birkhäuser Boston, Inc., 1985.

11. H.W. Lenstra, Jr., "Primality Test," in: H.W. Lenstra, Jr., R. Tijdeman (eds), *Computational Methods in Number Theory*, Math. Centre Trace **154/155**, Mathematisch Centrum, Amsterdam 1982, pp. 55-77.

12. S.M. Jennings, "A Special Class of Binary Sequences," University of London, 1980, Ph.D. Thesis.

13. Kencheng Zeng, C.H. Yang, and T.R.N. Rao, "On the Linear Consistency Test (LCT) in Cryptanalysis with Applications," presented to the Ninth Annual Crypto Conference, Santa Barbara, California, August 20-27, 1989. To appear in: Advances in Cryptology, Proc. of Crypto'89 (Lecture Notes in Computer Science), Springer-Verlag.

14. S. Mund, D. Gollmann, and T. Beth, "Some Remarks on the Cross Correlation Analysis of Pseudorandom Generators," *EUROCRYPT 87*, 1987, pp. 25-35.