# WHEN SHIFT REGISTERS CLOCK THEMSELVES

Rainer A. Rueppel

Crypto AG
6312 Steinhausen
Switzerland

Abstract:

A new class of sequences, which we term [d,k] self-decimated sequences, is investigated. For appropriate choices of [d,k] these sequences possess large periods, balanced k-distributions, large linear complexities, and moderate out-of-phase autocorrelation magnitudes. Furthermore, they are easy to generate. These properties suggest that [d,k] self-decimated sequences may have some applications in cryptography and spread spectrum communication.

## 1 INTRODUCTION

Imagine we let the output sequence of a binary linear feedback shift register (LFSR) determine its own clock in the following way: whenever the output symbol is a '0', d clock pulses are applied to the LFSR, and, in case the output symbol is a '1', k clock pulses are applied to the LFSR. Figure 1 illustrates the system.

D. Chaum and W.L. Price (Eds.): Advances in Cryptology - EUROCRYPT '87, LNCS 304, pp. 53-64, 1988.

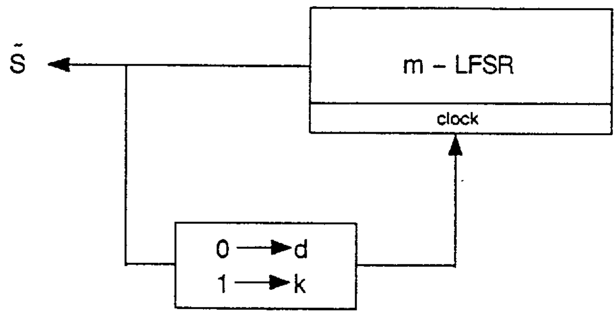


Fig. 1. Self-clocking LFSR

Suppose the above LFSR has a primitive connection polynomial $C(D) = 1 + D + D^2 + D^3 + D^4 + D^5$ and is started in state [1 1 1 1 1]. When the self-clocking rule [d,k] is chosen to be [1,2] (i.e., for a '0', the LFSR is clocked once, and for a '1' the LFSR is clocked twice), then the following periodic sequence will appear at the output of the system:

$$\tilde{s} = (11101010000110110011)_0^\infty$$

This sequence has remarkable properties: (1) the distribution of k-tuples is 'balanced' (to be more precise: for $1 \leq k \leq 3$, the frequencies of k-tuples differ by at most 2); (2) the linear complexity (or linear span) of $\tilde{s}$ is 20, which is the maximum possible for a sequence of period 20; (3) the periodic autocorrelation function of $\tilde{s}$ has a peak out-of-phase magnitude of 0. This self-clocking operation can be interpreted as a generalization of the well-known and widely-studied decimation operation for LFSR-sequences ([1],[2],[3],[4]). The conventional decimation of a sequence $s$ by a constant d is defined as the extraction of every d-th digit of $s$, usually denoted as $s$[d]. When a binary sequence $s$ is [d,k]-"self-clocked", then it is no longer decimated by a constant but by a function of the previous sequence digit; we will term the resulting sequence a <u>[d,k] self-decimated sequence</u>. Let $s$ be the original m-sequence produced by the LFSR in Figure 1. Then the following example compares ordinary decimation by 2, and [1,2] self-decimation of $r$ into $s$.

$$r = 1111100100110000101101 0100\ldots$$

$$r[2] = 1110010011000\ldots$$

$$r[1,2] = 111010100001101 10\ldots$$

Throughout this paper we will restrict ourselves to the case where the original sequences are maximum-length sequences over GF(2). Furthermore, without loss of generality, it is assumed that $0 < d,k < 2^L-1$, since, as with ordinary decimation, any d or k greater than $2^L-1$ can be reduced mod $2^L-1$.

If d is a unit mod $2^L-1$, (i.e. d has an inverse mod $2^L-1$), then

$$r[d,k] = r[d][1,k']$$

where

$$k' = k \cdot d^{-1} \mod 2^L - 1$$

That is, the self-decimation operation can be broken up into a constant decimation by d followed by a self-decimation of the special form [1,k']. If d is a unit mod $2^L-1$ then $r[d]$ is again an m-sequence of same degree and period. It is to be expected that for given self-clocking rules [d,k] certain properties like period or bit distribution are invariant over the set of all m-sequences of same degree. In general, there are $\Phi(2^L-1)\cdot(2^L-2)$ pairs [d,k] with d being a unit; $\Phi(n)$ denotes Euler's totient function. If $2^L-1$ is a Mersenne prime then all pairs [d,k] can be reduced to [d][1,k']. Thus, almost all cases can be covered by investigating self-decimation rules of the form [1,k].

Clearly the state diagram of the self-decimated m-LFSR will contain (one or more) cycles and tails. Depending on the initial state there may be a preperiod in the self-decimated sequence. A

<u>resetting sequence</u> is a subsequence of the original m-sequence which guarantees that the digit directly following the subsequence belongs to the self-decimated sequence. As an example let [d,k] be [1,2], then 0 is a resetting sequence. For, if the 0 itself belongs to the self-decimated sequence, so must its successor by the fact d=1; if, on the other hand, the 0 does not belong to the self-decimated sequence, its successor must by the fact k=2. This implies that, if a resetting sequence can be identified in the original m-sequence, then there exists only one cycle in the state diagram of the self-decimated m-LFSR, or equivalently, there exists only one self-decimated sequence (disregarding the preperiods for the moment).

Let us hypothetically assume that the original sequence is not an m-sequence but is comprised of N random bits which are repeated periodically. For [d,k] = [1,2], in the average every 1.5th digit is selected for the self-decimated sequence, or in other words, 2/3 of the original N random bits will appear in the self-decimated sequence. The first 0 among the N random bits will be a resetting subsequence. Thus, the period of the self-decimated sequence is expected to be approximately 2/3 N. As we will see in section 2 this is in perfect agreement with the theoretical results obtained for m-sequences. In section 3. some experimental data about characteristic properties like linear complexity and autocorrelation is shown.

## 2 THEORETICAL RESULTS

This section shall serve to demonstrate that despite the highly nonlinear setup of a self-clocking LFSR some analytical results can be obtained. The first topic of interest is the period.

*<u>Theorem 1</u>*: A [d,k]-self-decimated m-sequence of degree L has period

$$T_L = \left\lfloor \frac{2}{3}(2^L - 1) \right\rfloor$$

if [d,k] = g[1,2] mod $2^L-1$ with gcd(g,$2^L-1$) = 1.

<u>Proof:</u> 0 is a resetting sequence; any following digit in the original m-sequence must belong to the periodic part of the self-decimated sequence.

It follows that for any subsequence $01^m x$ of the m-sequence the digit x will belong to the self-decimated sequence if and only if m is even.

From the theory of m-sequences the frequencies of such subsequences are known:

$$\#\{0x\} = 2^{L-1} - 1$$
$$\#\{01^m x\} = 2^{L-m-1} \qquad m = 1,\dots,L-1$$
$$\#\{01^L x\} = 1$$

Thus, the number of digits that will appear in the periodic part of the self-decimated sequence can be found by a simple counting argument.

If L is odd we have

$$T_L = 2^{L-1} - 1 + 2^{L-3} + 2^{L-5} + \dots + 2^0$$

If L is even we have

$$T_L = 2^{L-1} - 1 + 2^{L-3} + 2^{L-5} + \dots + 2^1 + 1$$

Using the fact that

$$\sum_{i=0}^{n-1} q^i = \frac{q^n - 1}{q - 1}$$

we obtain

$$T_L = \frac{2}{3}\left(2^L - 2\right) \qquad \text{L odd}$$

$$T_L = \frac{2}{3}\left(2^L - 1\right) \qquad \text{L even}$$

which proves the theorem.

The identical argument can be carried out for $[d,k] = g[1,2^{L-1}]$ since $2[1,2^{L-1}] \equiv [2,1] \bmod 2^L-1$. In this case 1 is a resetting sequence and

$$T_L = \left\lceil \frac{2}{3}(2^L - 1) \right\rceil$$

Note that for odd L $[1,2^{L-1}]$ self-decimation yields a period which is one digit larger than for [1,2] self-decimation. The reason lies in the fact that the number of ones in an m-sequence exceeds the number of zeros by one.

*Theorem 2*: The absolute frequency of ones, $N_L(1)$, in the periodic part of a [d,k] self-decimated m-sequence of degree L is given as

$$N_L(1) = \left\lceil \frac{1}{3}(2^L - 1) \right\rceil$$

if $[d,k] = g[1,2] \bmod 2^L-1$ with $\gcd(g,2^L-1) = 1$.

Proof: 0 is a resetting sequence; any following 1 in the original m-sequence must belong to the periodic part of the self-decimated sequence.

Thus, for any subsequence $01^{m+1}$ of the m-sequence the final 1 will belong to the self-decimated sequence if and only if m is even.

From the theory of m-sequences the frequencies of such subsequences are known:

$$\#\{01^{m+1}\} = 2^{L-m-2} \qquad m = 0, \ldots, L-2$$
$$\#\{01^L\} = 1$$

If L is odd we have

$$N_L(1) = 2^{L-2} + 2^{L-4} + \ldots + 2^1 + 1$$

If L is even we have

$$N_L(1) = 2^{L-2} + 2^{L-4} + \ldots + 2^0$$

We obtain

$$N_L(1) = \frac{1}{3}\left(2^L + 1\right) \qquad \text{L odd}$$

$$N_L(1) = \frac{1}{3}\left(2^L - 1\right) \qquad \text{L even}$$

which proves the theorem.

Note that for even L the bit distribution is perfectly balanced, i.e., $N_L(0) = N_L(1) = T_L/2$.

*Theorem 3*: Let [d,k] = g[1,2] mod $2^L-1$ with gcd(g,$2^L-1$) = 1. Then the absolute frequencies of bit pairs, $N_L(b_1,b_2)$, within one period of a [d,k] self-decimated m-sequence of degree L are bound by

$$N_L^0 \le N_L(b_1,b_2) \le N_L^0 + 2$$

with

$$N_L^0(00) = \left\lfloor \frac{1}{6}\left(2^L - 1\right) \right\rfloor - 1$$

$$N_L^0(01) = \left\lfloor \frac{1}{6}\left(2^L - 1\right) \right\rfloor$$

$$N_L^0(10) = \left\lfloor \frac{1}{6}\left(2^L - 4\right) \right\rfloor$$

$$N_L^0(11) = \left\lfloor \frac{1}{6}\left(2^L - 4\right) \right\rfloor$$

Proof: *case a*: 00

0 is a resetting sequence; any following pair '00' in the original m-sequence must belong to the periodic part of the self-decimated sequence.

It follows that for any subsequence $01^m00$ of the m-sequence the pair 00 will belong to the self-decimated sequence if and only if m is even.

From the theory of m-sequences the frequencies of such subsequences are known:

$$\#\{000\} = 2^{L-3} - 1$$
$$\#\{01^m 00\} = 2^{L-m-3} \qquad m = 1,\dots,L-3$$

Subsequences longer than L may or may not exist as long as the number of consecutive 1's does not exceed L. Thus

$$\#\{01^m 00\} = 0 \text{ or } 1 \qquad m = L-2,\dots,L$$

It follows that

$$N_L(00) \;\geq -1 + \sum_{\substack{m=0 \\ m \text{ even}}}^{L-3} 2^{L-m-3} \;=\; -1 + \left\lfloor \frac{1}{6}\left(2^L - 1\right) \right\rfloor$$

which proves the lower bound.

From the uncertain overlong subsequences at most 2 could contribute an entry, which proves the upper bound.

*case b*: 01

For any subsequence $01^m01$ of the original m-sequence the final pair 01 will belong to the self-decimated sequence if and only if m is even.

$$\#\{01^m 01\} = 2^{L-m-3} \qquad m = 0,\dots,L-3$$
$$\#\{01^m 01\} = 0 \text{ or } 1 \qquad m = L-2,\dots,L$$

It follows that

$$N_L(01) \;\geq \sum_{\substack{m=0 \\ m \text{ even}}}^{L-3} 2^{L-m-3} \;=\; \left\lfloor \frac{1}{6}\left(2^L - 1\right) \right\rfloor$$

*case c*: 10

For any subsequence $01^mx0$ (x arbitrary) of the original m-sequence the final pair 10 (x dropped) will belong to the self-decimated sequence if and only if m is odd.

$$\#\{01^m x0\} = 2^{L-m-2} \qquad m = 1,\dots,L-3$$
$$\#\{01^m x0\} = 0 \text{ or } 1 \qquad m = L-2,\dots,L$$

Therefore

$$N_l(10) \geq \sum_{\substack{m=1 \\ m\,odd}}^{l-3} 2^{l-m-2} = \left\lfloor \frac{1}{6}(2^l-4) \right\rfloor$$

*case d*: 11
This case is analogous to case c with the final 0 in the subsequence replaced by 1.

Since the lower bounds differ by at most 1, theorem 3 implies that the absolute frequencies of bit pairs cannot differ by more than 3.

The trace from $GF(2^L)$ into $GF(2)$ is defined as

$$Tr(\beta) = \beta + \beta^2 + \ldots + \beta^{2^{L-1}}$$

where $\beta$ is an element of $GF(2^L)$. With the help of the trace function the jth digit of an m-sequence can be compactly expressed [4] as

$$r_j = Tr\left(A\alpha^j\right)$$

where $\alpha$ is a root of the minimal polynomial of $r$, and A relates to the initial phase of $r$.

For a [d,d+1] self-decimated m-sequence we obtain

$$s_j = Tr\left(A\alpha^{e_j}\right)$$

with

$$e_j = d \cdot j + \sum_{k=0}^{j-1} s_k$$

This leads to the following (nonlinear) recursion of the exponents

$$e_{j+1} = e_j + d + Tr\left(A\alpha^{e_j}\right)$$

As was mentioned before, the state diagram of a [d,k] self-decimated m-LFSR contains (one or more) cycles and tails. Every tail must contain an initial state (which we call a root in this context), and must finally join either another tail or a cycle. These junction states are particular in the sense that they have two predecessors but only one successor. This implies that for a junction state's exponent $e_{j+1}$ there exist two distinct exponents $e_j$ and $e_j'$, corresponding to the two distinct predecessors. Consequently

$$e_j - e'_j = Tr\left(A\alpha^{e'_j}\right) - Tr\left(A\alpha^{e_j}\right)$$

Without loss of generality assume $e_j > e_j'$. Then, for [d,d+1] self-decimation, we obtain

$$1 = Tr\left(A\alpha^{e_j - 1}\right) - Tr\left(A\alpha^{e_j}\right)$$

This equation tells us that a transition from 1 to 0 has occurred in the original m-sequence. The number of such transitions is $2^{L-2}$. This proves the following theorem.

*Theorem 4:* The number of roots, (i.e. states with no predecessor) in the state diagram of a [d,d+1] self-decimated m-LFSR of length L is

$$N_L(\text{roots}) = 2^{L-2}$$

Since roots cannot be part of a cycle the following corollary is obvious.

*Corollary 5:* The period of a [d,d+1] self-decimated m-sequence of degree L is bound from above by

$$T_L \leq 2^L - 1 - 2^{L-2} = \left\lfloor \frac{3}{4}\left(2^L - 1\right) \right\rfloor$$

## 3 EXPERIMENTAL RESULTS

Extensive simulations have been run for [1,2] self-decimated m-sequences of degrees L=3,...,11. They showed that, for given L, also the pair distributions, (beside period and bit distributions), were independent of the minimal polynomial of the m-sequence (see table 1).

| $L$ | $T_L$ | $N_L(0)$ | $N_L(1)$ | $N_L(00)$ | $N_L(01)$ | $N_L(10)$ | $N_L(11)$ |
|---|---|---|---|---|---|---|---|
| 4 | 10 | 5 | 5 | 2 | 3 | 3 | 2 |
| 5 | 20 | 9 | 11 | 4 | 5 | 5 | 6 |
| 6 | 42 | 21 | 21 | 10 | 11 | 11 | 10 |
| 7 | 84 | 41 | 43 | 20 | 21 | 21 | 21 |

Table 1. Periods, bit, and pair distributions.

Exhaustive searches over all primitive polynomials of degree L = 5,6,7,8 revealed the following agerages and minimum values for the linear complexities of [1,2] self-decimated m-sequences:

| $L$ | $T_L$ | $L_{avg}$ | $L_{min}$ |
|---|---|---|---|
| 5 | 20 | 19,3 | 16 |
| 6 | 42 | 38,7 | 33 |
| 7 | 84 | 82 | 78 |
| 8 | 170 | 169,3 | 166 |

Table 2. Linear complexities

The proximity of $L_{avg}$ to the period length $T_L$ and the largeness of the minimal encountered linear complexity $L_{min}$ speak for themselves.

Another topic of interest is the periodic autocorrelation function. Exhaustive searches over all primitive polynomials of degrees L = 4,5,6,7 revealed the following averages $R_{avg}$ and minimum values $R_{min}$ for the peak out-of-phase autocorrelation magnitude of [1,2] self-decimated sequences:

| $L$ | $R_{avg}$ | $R_{min}$ |
|---|---|---|
| 4 | 4 | 2 |
| 5 | 4 | 0 |
| 6 | 9.3 | 6 |
| 7 | 20.5 | 12 |

Table 3. Out-of-phase autocorrelation magnitudes

## 4 CONCLUSION

[d,k] self-decimated m-sequences are (almost) as easy to generate as m-sequences; for appropriately chosen [d,k] they exhibit similar properties as m-sequences with respect to period, k-distributions, and autocorrelation. But they behave much more like 'truly' random sequences as is indicated by the high linear complexity values. Therefore [d,k] self-decimated sequences may have some applications in cryptography and spread spectrum communication.
But a word of caution has to be added; if a [d,k] self-decimated m-LFSR is employed alone and [d,k] are made public, then from its output sequence the feedback polynomial and the initial state of the LFSR are easily retrieved (a system of linear equations has to be solved).

## 5 REFERENCES

[1] N. Zierler, "Linear recurring sequences", J. Soc.. Indust. Appl. Math., Vol. 7, 1959.

[2] D.V. Sarwate, M.B. Pursley, "Crosscorrelation properties of pseudorandom and related sequences", Proc. of the IEEE, Vol. 68, May 1980.

[3] J.L. Massey, R.A. Rueppel, "Linear ciphers and random sequence generators with multiple clocks", Proc. of Eurocrypt 84, Paris, Lecture Notes in Computer Science, Vol. 209, Springer-Verlag.

[4] R.A. Rueppel, "Analysis and design of stream ciphers", Springer-Verlag, 1986.