PSEUDO-RANDOM SEQUENCE GENERATORS USING STRUCTURED NOISE

R. S. Safavi-Naini and J. R. Seberry

Stream ciphers use the output of a Pseudo-Random (*PR*) generator to mask the information stream. The security of these cipher systems ultimately depends on the structure of the *PR* generator. There are some minimum necessary criteria such as long period, flat statistical distribution and high linear complexity that the *PR* generator of a stream cipher system should satisfy to resist the basic cryptanalytic attacks on such systems. We propose a class of *PR* generators using the coset elements of a Reed-Muller code. The linear complexity of these generators is analysed and conditions that assure the highest possible linear complexity for them are specified. It is shown that the above mentioned criteria do not gurantee the security of a stream cipher system and the proposed *PR* generator, although it satisfies all of them, is not secure.

1. Introduction.

Stream ciphers assimilate the one time pad, the only provably perfect secure system. However with the replacement of the random generator by a pseudo-random (*PR*) one, the perfect security of the system vanishes. It is easy to see that the assessment of the security of these systems is directly related to the properties of the *PR* generator. There are some necessary criteria which must be satisfied by the *PR* generator of a secure stream cipher. It is recognised that these generators should satisfy Golomb's criteria and have high linear complexity [1], [3]. Linear feedback shift register (*LFSR*) generated sequences satisfy Golomb's criteria but have a small linear complexity and hence fail to offer high security. There have been many attempts to devise algorithms based on *LFSR*'s that retain the good properties of these sequences and increase the linear complexity of them. In this paper we propose noise addition as a mechanism to achieve the above-mentioned goal. In the rest of this section we briefly review some relevant background material. In Section 2, the properties of some special subsets of the binary vector space of dimension $2^m - 1$ are studied and the results are used in Section 3 to develop a new class of stream ciphers. The paper is concluded by establishing the insufficiency of linear complexity and Golomb's criteria for the security of *PR* generators.

Linear Equivalence. A linear feedback shift register (*LFSR*) is a finite state machine, the state of which at time t is determined by its content at time t and its next state is determined by its feedback function (we consider the binary case only).

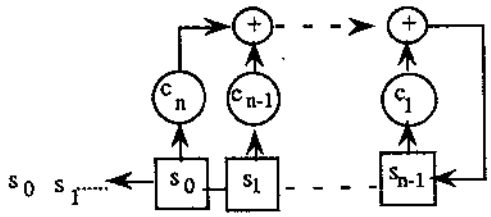

Figure 1.

The feedback polynomial of the *LFSR* is defined as

$$C(D) = 1 + c_1 D + c_2 D^2 + \ldots + c_m D^m,$$

where c_i's are binary feedback coefficients. Let $s = s_0, s_1, s_2, \ldots$ denote the semi-infinite output sequence of the *LFSR*. The D-transform of s is $S(D)$ defined by

$$S(D) = s_0 + s_1 D + s_2 D^2 + \ldots .$$

The output sequence s is a periodic sequence, the period of which is determined by the properties of $C(D)$ [1]. An m-sequence has the maximum period, $2^m - 1$, and corresponds to the case when the feedback polynomial is primitive. (See the contribution by Dawson in these Proceedings.) The D-transform of a sequence s generated by an *LFSR* can be written as:

$$S(D) = \frac{P(D)}{C(D)},$$

where P is a polynomial with degree less than m. In fact there is a one-to-one correspondence between polynomials of degree less than m and the set of $2^m - 1$ output sequences of the *LFSR* corresponding to the $2^m - 1$ non-zero possible initial conditions.

Every periodic sequence s of period T can be thought of as generated by a *LFSR* of length T. The D-transform of the sequence can be written as:

$$S(D) = \frac{S(D)^*}{1 + D^T}$$

where $S(D)^*$ denotes the D-transform of the first period. By cancelling the common factors from the numerator and the denominator one can find the *LFSR* of minimum length that can generate the sequence. The length of this *LFSR* is the linear equivalence of the sequence and is taken as a measure of its complexity.

The first periods of periodic sequences of period T constitute a T-dimensional vector space V_T. The linear equivalence of $\mathbf{v}$ in V_T is denoted by $L(\mathbf{v})$ and defined as the linear equivalence of the semi-infinite sequence with first period equal to $\mathbf{v}$. The D-transform of $\mathbf{v}$ is defined as the D-transform of the first period of this sequence. One set of basic vectors can be obtained by decomposing $1 + D^T$ into irreducible factors:

$$1 + D^T = \prod_i C_i(D),$$

where $C_i(D)$'s are irreducible polynomials over $GF(2)$. By applying a partial fraction expansion to $S(D)$, the D-transform of a sequence s of period T, we have

$$S(D) = \sum_i \frac{P_i(D)^*}{C_i(D)},$$

which shows the sequence can be generated by adding (modulo 2) the output of some basic *LFSR*'s with irreducible feedback polynomials given by $C_i(D)$ and their initial content determined by $P_i(D)$. The total length of the basic *LFSR*'s is equal to T.

Example 1: Let the first period of a sequence s be:

1 0 1 0 1 0 1 1 1 0 0 0 0 1 1

and $S(D)$ denotes its D-transform. Partial fraction expansion of $S(D)$ results in:

$$\begin{aligned} S(D) &= \frac{1 + D^2 + D^4 + D^6 + D^7 + D^8 + D^{13} + D^{14}}{1 + D^{15}} \\ &= \frac{D}{1 + D + D^2} + \frac{1}{1 + D + D^2 + D^3 + D^4} \end{aligned}$$

So the sequence can be generated by a *LFSR* of length 6.

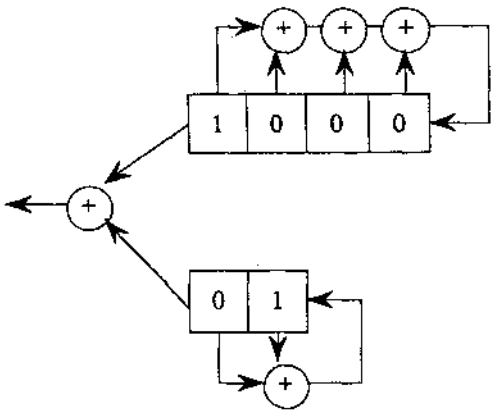

Figure 2.

Reed-Muller Code. Binary Reed-Muller code of order r and length $2^m - 1$, denoted by $RM(r, m)$ is the set of all Boolean functions f of m Boolean variables $v_1, v_2, \ldots, v_m$ where $f(v_1, v_2, \ldots, v_m)$ is a polynomial of order at most r when written in algebraic normal form [2]. The Reed-Muller code of order one comprises the set of linear Boolean functions of m variables and can be partitioned into two subsets one of which is the set of complement vectors of the other:

$$RM(1, m) = O_m \cup (1 + O_m),$$

where O_m is a subspace by itself and has an all-zero column. The linear code obtained by deleting this column, denoted by O_m^*, is in fact the set of m-sequences generated by a *LFSR* with the primitive feedback polynomial equal to the parity check polynomial of O_m^*. The minimum distance of O_m^* is 2^{m-1} and can correct $2^{m-2}-1$ errors. Hence in the standard array correspnding to O_m^*, all vectors of weight $2^{m-2}-1$ appear as coset leaders and these cosets can be identified by their leaders [3].

Example 2: The generator polynomial $g(x)$ and the parity check polynomial $h(x)$ of O_3^* are :

$$g(x) = 1 + x^2 + x^3 + x^4 = (1+x)(1+x+x^3)$$
$$h(x) = 1 + x^2 + x^3$$

The non-zero codewords (listed below) can be generated by a *LFSR* with feedback polynomial $C(D) = 1 + D^2 + D^3$:

1	0	0	1	0	1	1
1	1	0	0	1	0	1
1	1	1	0	0	1	0
0	1	0	1	1	1	0
0	1	1	1	0	0	1
0	0	1	0	1	1	1
1	0	1	1	1	0	0

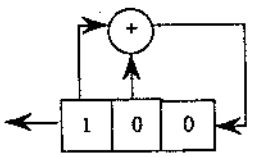

Figure 3.

2. PR Sequences from RM(1, m).

The following propositions give some classification of the elements of V_T in terms of their linear equivalence.

Proposition 1. *The linear equivalence of the vectors of O_m^* is m.*

Proof. Follows from the fact that the codewords are m-sequences.

Proposition 2. *Let $\mathbf{e} \in V_T$ ($T = 2^m - 1$), be a coset leader of O_m^* with $L(\mathbf{e}) = h$. The linear equivalence $l_{\mathbf{e}}$ of all but at most one of the elements of the coset with $\mathbf{e}$ as the leader is the same. The only possible values of $l_{\mathbf{e}}$ are h and $h \pm m$.*

Proof. See [4].

So cosets of O_m^* are subsets of constant (almost) linear equivalence and coset decomposition of V_T can be regarded as decomposition in terms of linear complexity also. In fact, to specify the linear equivalence of a vector, it is almost always enough to determine the coset to which it belongs. This is the standard decoding problem of O_m^* which is well studied [2].

Proposition 3. *Let* $\mathbf{e} \in V_T$ $(T = 2^m - 1)$ *and* $w(\mathbf{e}) = 1$. *Then* $L(\mathbf{e}) = 2^m - 1$.

Proof. See [4].

Corollary. *In every coset of weight one of* O_m^* *there exists exactly one vector of linear equivalence* $2^m - 1 - m$. *All the other vectors have linear equivalence* $2^m - 1$.

This gives an explicit expression of the linear equivalence of all vectors of cosets of weight one and shows that almost all these vectors have the highest linear equivalence. Dependence of the linear equivalence of coset elements on the weight of the coset leaders cannot be extended to higher weights for which the specific error pattern affects the complexity of the vector. Let $w(\mathbf{e})$ denote the Hamming weight of $\mathbf{e}$.

Proposition 4. *The linear equivalence of a vector* $\mathbf{e} \in V_T$ *with* $w(\mathbf{e}) = 2$, *depends on the distance between the two non-zero components. If this distance is denoted by* t *and* $\deg(\gcd(1 + D^t, 1 + D^T)) = u$, *then* $L(\mathbf{e}) = T - u$.

Proof. See [4].

Corollary. *The linear equivalence of vectors of cosets of weight two of* O_m^* *depends on the distance* t *between the two non-zero components of the coset leader. If we have* $T = 2^m - 1$ *and* $\deg(\gcd(1 + D^t, 1 + D^T)) = u$, *then the linear complexity of the coset element is at least* $2^m - 1 - u - m$.

Example 3. Consider the vectors of weight 2 in V_{15}. Let t be the distance between the non-zero bits of a vector $\mathbf{e}$ where $w(\mathbf{e}) = 2$. The possible values of t are $1, 2, \ldots, 14$ which result in the following values for the linear equivalence:

t	1	2	3	4	5	6	7	8	9	10	11	12	13	14
$L(\mathbf{e})$	14	14	12	14	10	12	14	14	12	10	14	12	14	14

It should be noted that the distance between the two non-zero components and not their actual position is important. So all the vectors listed below have the same linear equivalence 10 corresponding to $t = 10$ in the above table:

$\mathbf{e}_1$	1 0 0 0 0 0 0 0 0 0 1 0 0 0 0
$\mathbf{e}_2$	0 1 0 0 0 0 0 0 0 0 0 1 0 0 0
$\mathbf{e}_3$	0 0 1 0 0 0 0 0 0 0 0 0 1 0 0
$\mathbf{e}_4$	0 0 0 1 0 0 0 0 0 0 0 0 0 1 0
$\mathbf{e}_5$	0 0 0 0 1 0 0 0 0 0 0 0 0 0 1

For cosets of weight greater than two the linear complexity is independent of the shift in the error pattern.

Proposition 5. *Let* $E^*(D) = D^k E(D)$, $E(0) = 1$, *denote the* D*-transform of a coset leader* $\mathbf{e}$ *and* $\deg(\gcd(E(D), 1 + D^T)) = u$ $(T = 2^m - 1)$. *The linear complexity of a coset element is at least* $2^m - 1 - u - m$.

Proof. The D-transform of a sequence with first period given by $\mathbf{e}$ is

$$\frac{E(D)}{1 + D^T} \quad (T = 2^m - 1).$$

It follows that the linear equivalence of this sequence is $T-u$ and the linear equivalence of a vector of the coset with $\mathbf{e}$ as a leader is at least $T - u - m$. In fact there is exactly one vector of this complexity and the linear complexity of the rest is $T - u$.

Corollary 1. *Let* $\mathbf{e} \in V_T$ *and* $E(D)$ *be a power of a prime polynomial. Then* $L(\mathbf{e}) = 2^m - 1$.

Corollary 2. *Let* $\mathbf{e}$ *and* $\mathbf{e}'$ *be two vectors of length* $T = 2^m - 1$ *such that*

$$E_{\mathbf{e}}(D) = D^T E_{\mathbf{e}'}(D^{-1}).$$

Then $L(\mathbf{e}) = L(\mathbf{e}')$.

Proof. Follows from the fact that

$$\gcd(E_{\mathbf{e}}(D), 1 + D^T) = \gcd(E_{\mathbf{e}'}(D), 1 + D^T).$$

3. Key Generators Using Structured Noise.

One can use noise addition to increase the linear equivalence of a vector. A *PR*-generator based on this idea is shown in Figure 4.

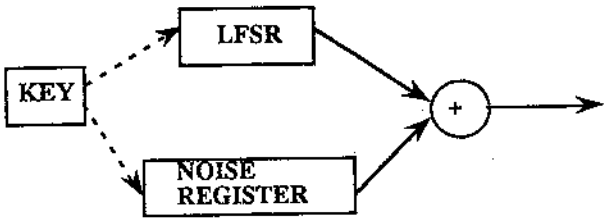

Figure 4.

The *LFSR* generates an m-sequence of period T which is added modulo two to a noise sequence of the same period. The noise register contains a class of noise vectors of length T that have high linear equivalence and are easy to generate (easy to describe). The class is parametrized by a piece of the key information. A common key

in transmitter and receiver produces the same *PR*-sequence at both ends. This is an easy and fast way of enhancing the linear complexity of an m-sequence. A large number of noise sequences stored in the register will prevent a cryptanalyst from recovering the m-sequence easily. Propositions 3 and 4 imply that all vectors of weight one and a large number of vectors of weight two (the linear equivalence of the vector should be determined) can be included in the noise register. Using Proposition 5 it can be seen that finding the proper noise vector is equivalent to finding polynomials which are relatively prime to $1+D^T$, or have a gcd of small degree. Proposition 6 specifies a class of polynomials that correspond to vectors of high linear complexity. This result can be combined with the corollary to Proposition 5 to characterise suitable noise vectors. We need the following definition.

Definition ([3]). The *cyclotomic coset* mod n over $GF(q)$ which contains s is

$$C_s = \{s, sq, sq^2, ..., sq^{T_s}\}, \text{ where } sq^{m_s} \equiv s \bmod n,\ T_s = m_s - 1.$$

Proposition 6. *A sequence of length* $2^m - 1$ *that consists of* t *consecutive ones, where* t *is a prime such that* $\gcd(t, 2^m - 1) = 1$ *and* t *has only two cyclotomic cosets, has linear equivalence* $2^m - 1$.

Proof. Let $T = 2^m - 1$. Since, by hypothesis, $\gcd(t, 2^m - 1) = 1$, it follows that $\gcd(1 + D^t, 1 + D^T) = 1$. Further, since t has only two cyclotomic cosets,

$$1 + D^t = (1 + D) \cdot \sum_{i=0}^{t-1} D^i.$$

Now it is easy to see that

$$\gcd\left(1 + D^{2^m-1}, \sum_{i=0}^{t-1} D^i\right) = 1,$$

where the second polynomial corresponds to a vector of t consecutive zeros.

Security. It is easy to see that the *PR* generator discussed in this section satisfies Golomb's criteria [1] closely, if we restrict the weight of the noise sequences used in the noise register to small values (compared to the length of the *LFSR*). It is also noted that the proper selection of noise vectors ensures a lower bound on the linear complexity of the output sequence. However the small weight of the noise vectors implies that the first period of the sequence can be approximated by a *LFSR* sequence and would be highly predictable after $2m$ (m is the length of *LFSR*) bits of it were intercepted. So the sequence is not secure.

4. Concluding Remarks.

The study of linear complexity of a Reed-Muller code and its cosets suggests a new way of increasing the linear equivalence of an m-sequence. The results are applied to devise a *PR* generator. The added complexity is the result of adding a noise vector

with high linear equivalence to the m-sequence. If the weight of the noise vector is one, the statistical parameters of the original m-sequence would not be greatly affected and the resulting sequence would closely satisfy the known criteria of security. It is noticed that in general for noise vectors of small weight the generator is not secure because the output can be approximated by an m-sequence. On the other hand, noise vectors of higher weight deteriorate the statistical properties of the m-sequence.

The PR sequences obtained from the elements of cosets of weight one demonstrate the insufficiency of the criteria used for assessing the PR generators of stream ciphers. Providing criteria to measure the security of such systems remains an open problem.

References

1. H. Becker and F. Piper, *Cipher Systems.* (North Books, London, 1982.)
2. F. J. MacWilliams and N. J. Sloane, *The Theory of Error-Correcting Codes.* (North-Holland Publishing Company, 1978.)
3. R. A. Rueppel, *Analysis and Design of Stream Ciphers.* (Springer-Verlag, Berlin 1986.)
4. R. S. Safavi-Naini and J. R. Seberry, 'Pseudo-Random Sequences from Codes' Technical Report CS89/10, University College, University of New South Wales Canberra, Australia.

Department of Computer Science, University College, The University of New South Wales, Australian Defence Force Academy, Canberra, ACT 2600, AUSTRALIA.