A Fourier Transform Approach to the Linear Complexity of Nonlinearly Filtered Sequences

James L. Massey and Shirlei Serconek*
Signal and Information Processing Laboratory
Swiss Federal Institute of Technology
ETH-Zentrum
CH-8092 Zürich, Switzerland

Abstract

A method for analyzing the linear complexity of nonlinear filterings of PN-sequences that is based on the Discrete Fourier Transform is presented. The method makes use of "Blahut's theorem", which relates the linear complexity of an N-periodic sequence in $GF(q)^N$ and the Hamming weight of its frequency-domain associate. To illustrate the power of this approach, simple proofs are given of Key's bound on linear complexity and of a generalization of a condition of Groth and Key for which equality holds in this bound.

1 Introduction

Fourier transforms in a Galois field play an important role in the study and processing of $GF(q)$-valued signals, particularly in coding theory. By revisiting many topics by way of the *frequency-domain*, deeper understanding and alternative encoding and decoding techniques can be found (Blahut [1]).

By exploiting "Blahut's theorem", which states that the linear complexity of an N-periodic sequence in $GF(q)^N$ and the Hamming weight of its *frequency-domain* associate are equal, we use Discrete Fourier Transform (DFT) techniques here to study the linear complexity of nonlinear filterings of PN(pseudo-noise)-sequences. To illustrate the power of this approach, we give a simple and transparent proof of a generalization of a result of Groth [2] and Key [3].

Groth applied second-order boolean functions to the stages of an LFSR with a primitive connection polynomial. No stage was allowed to be used more than once. But Groth, as well as Key [3]. as pointed out by Rueppel [7], "limited

*This work was done while the author was on leave from CEPESC, Cx Postal 02976, Brasília, DF, BRASIL, CEP 70610-200

Y.G. Desmedt (Ed.): Advances in Cryptology - CRYPTO '94, LNCS 839, pp. 332-340, 1994.

himself to consider only those sequences which are available at the stages of an LFSR, in particular he allowed only phase differences of at most the length of LFSR. But when the speed of the LFSR is taken as an additional parameter, this theory must be extended to allow arbitrary phase differences. But then even the result of Key's which might be considered as his most solid one, namely, that a 2nd order product of 2 distinct phases of the same sequence never degenerates, is no longer true." Using the DFT method that we develop in section 4, we show that Key's non-degeneration result always holds when the length of the LFSR is a prime; the restrictions imposed by Groth and Key on the phase differences are unnecessary in this case.

We will discuss only the binary case $q = 2$ in this paper, but the Discrete Fourier Transform properties hold in the general case and all our results readily generalize to any finite field.

2 Period and Linear Complexity of Sequences

The linear complexity, $\mathcal{L}(\tilde{s})$, of the sequence $\tilde{s} = s_0, s_1, \cdots$, $s_i \in F$, F an arbitrary field, is the length L of the shortest linear feedback shift-register (LFSR) that can generate $\tilde{s}$ when the first L digits of $\tilde{s}$ are initially loaded in the register. Equivalently, the linear complexity is defined to be the smallest nonnegative integer L such that there exist coefficients $c_1, c_2, \cdots, c_L$ in F such that

$$s_j + c_1 s_{j-1} + \cdots + c_L s_{j-L} = 0 \,, \qquad j \geq L \,.$$

Linear complexity is very useful in the study of stream ciphers; a necessary condition for security of a running-key generator is that it produce a sequence with large linear complexity.

The sequence $\tilde{s}$ will be called N-periodic if N is a positive integer such that $s_i = s_{i+N}$ for all $i \geq 0$. If we characterize a periodic sequence $\tilde{s} = s_0, s_1, \cdots$ by its D-transform

$$S(D) = s_0 + s_1 D + s_2 D^2 + \cdots \,,$$

then it is well known [5] that

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

where

$$C(D) = 1 + c_1 D + c_2 D^2 + \cdots + c_L D^L \,, \qquad c_L \neq 0$$

and

$$\deg P(D) < \deg C(D) \,.$$

The polynomial $C(D)$ is the connection polynomial of an LFSR of length L that generates $\tilde{s}$ when its initial state is $[s_0, s_1, \cdots, s_{L-1}]$. If $\gcd(P(D), C(D)) = 1$, then this is the unique shortest LFSR that can generate the sequence $\tilde{s}$.

3 The Discrete Fourier Transform (DFT)

3.1 Definition of the DFT

We write $\mathbf{a}^N$ to denote the N-tuple $[a_0, a_1, \cdots, a_{N-1}] \in F^N$ where F is an arbitrary field. Under the assumption that there exists a primitive N-th root of unity ω in the field F, the Discrete Fourier Transform (DFT) of a *time-domain* N-tuple $\mathbf{a}^N = [a_0, a_1, \cdots, a_{N-1}]$ is defined to be the *frequency-domain* N-tuple $\mathbf{A}^N = [A_0, A_1, \cdots, A_{N-1}]$, where the components of $\mathbf{A}^N$ are given by

$$A_j = \sum_{i=0}^{N-1} a_i \omega^{ij} \quad for\ j = 0, 1, \cdots N-1.$$

The sequence $\mathbf{a}^N$ can be recovered from $\mathbf{A}^N$ by the inverse DFT in the manner

$$a_j = \frac{1}{N^*} \sum_{i=0}^{N-1} A_i \omega^{-ij} \quad for\ j = 0, 1, \cdots N-1,$$

where

$$N^* = \begin{cases} N\ modulo\ p & \text{, if the characteristic of } F \text{ is a prime } p\ . \\ N & \text{, if the characteristic of } F \text{ is } 0\ . \end{cases}$$

The components of the N-tuples $\mathbf{a}^N$ and $\mathbf{A}^N$ can be used as coefficients to form two polynomials $a(X)$ and $A(X)$ in the indeterminant X whose degrees are at most $N-1$, namely

$$a(X) = a_0 + a_1 X + a_2 X^2 + \cdots + a_{N-1} X^{N-1}$$

$$A(X) = A_0 + A_1 X + A_2 X^2 + \cdots + A_{N-1} X^{N-1}\ .$$

In this polynomial representation, the DFT relations can be written simply as

$$A_j = a(\omega^j)\ \ and\ \ a_j = \frac{1}{N^*} A(\omega^{-j}) \quad for\ j = 0, 1, \cdots (N-1).$$

3.2 Properties of the DFT

The following properties of the DFT are valid in any field F containing a primitive N-th root of unity ω. Proofs may be found in [1]. A double parenthesis $(())$ about an integer denotes that this integer should be taken modulo N.

3.2.1 The Shifting Property

Assume that the time-domain N-tuple $\mathbf{b}^N = [b_0, b_1, \cdots, b_{N-1}]$ is formed by shifting the N-tuple $\mathbf{a}^N = [a_0, a_1, \cdots, a_{N-1}]$ cyclically to the *left* by k positions, i.e.,

$$b_i = a_{((i+k))} \quad for\ i = 0, 1, \cdots N-1,$$

then the components of the frequency-domain N-tuple $\mathbf{B}^N$ are given by

$$B_i = A_i.\omega^{-ki} \quad for\ i = 0, 1, \cdots N-1.$$

Equivalently,

$$B(X) = A(\omega^{-k}X)\ .$$

3.2.2 The Convolution Property

If the time-domain N-tuple $\mathbf{c}^N = [c_0, c_1, \cdots, c_{N-1}]$ is given by the component-wise product $c_i = a_i.b_i$ $(i = 0, 1, \cdots N-1)$ of N-tuples $\mathbf{a}^N$ and $\mathbf{b}^N$, then the frequency-domain N-tuple $\mathbf{C}^N = [C_0, C_1, \cdots, C_{N-1}]$ can be found by cyclically convolving the N-tuples $\mathbf{A}^N$ and $\mathbf{B}^N$ in the manner

$$C_j = \frac{1}{N^*}\sum_{k=0}^{N-1} A_{((j-k))}B_k \quad for\ j = 0, 1, \cdots N-1\ .$$

Equivalently,

$$C(X) = \frac{1}{N^*}A(X)B(X) \ \ mod\ X^N - 1.$$

3.2.3 The Conjugacy Constraints Property

Let the elements of the frequency-domain N-tuple $\mathbf{A}^N$ belong to the finite field $GF(q^m)$ where N divides $q^m - 1$, which condition is necessary and sufficient to ensure that $GF(q^m)$ contains primitive N-th roots of unity. Then the elements of the time-domain N-tuple $\mathbf{a}^N$ belong to the subfield $GF(q)$ if and only if the elements of the frequency-domain N-tuple satisfy the conjugacy constraints:

$$A_j^q = A_{((qj))} \quad 0 \le j < N\ .$$

Note that the set of indices $\{((qj)) \mid 0 \le j < N\}$ is a cyclotomic coset modulo N and that its cardinality must be a divisor of N.

3.2.4 Blahut's Theorem

Consider the frequency-domain vector $\mathbf{S}^N = [S_0, S_1, \cdots, S_{N-1}]$ associated with the time-domain vector $\mathbf{s}^N = [s_0, s_1, \cdots, s_{N-1}]$ by the DFT defined by a primitive N-th root of unity ω in F. Then, for $\omega = \alpha^{-1}$,

$$s_j = \frac{1}{N^*}\sum_{i=0}^{N-1} S_i\omega^{-ij} = \frac{1}{N^*}\sum_{i=0}^{N-1} S_i\alpha^{ij}$$

holds for all $j \ge 0$ since replacing j by $j + N$ leaves the sum unchanged. Thus the D-transform of $\tilde{s}$ is

$$\sum_{j=0}^{\infty} s_j D^j = \frac{1}{N^*}\sum_{j=0}^{\infty}\sum_{i=0}^{N-1} S_i\alpha^{ij} D^j$$

$$= \frac{1}{N^*} \sum_{i=0}^{N-1} S_i \sum_{j=0}^{\infty} (\alpha^i D)^j$$

$$= \frac{1}{N^*} \sum_{\substack{i=0 \\ S_i \neq 0}}^{N-1} S_i \frac{1}{1-\alpha^i D} .$$

Because the roots of the denominator polynomials are $\alpha^{-i} = \omega^i$ for those i such that $0 \leq i < N$ and $S_i \neq 0$, these roots are all distinct. Thus, letting $w(.)$ denote the Hamming weight (i.e., the number of nonzero components) of the enclosed N-tuple), we have

$$\sum_{j=0}^{\infty} s_j D^j = \frac{P(D)}{C(D)}$$

where

$$C(D) = \prod_{\substack{i=0 \\ S_i \neq 0}}^{N-1} (1 - \alpha^i D) \tag{1}$$

has degree $w(\mathbf{S}^N)$, where $\deg P(D) < \deg C(D)$ and $\gcd(P(D), C(D)) = 1$. By (1) and the last comment in section 2 we now have

$$\mathcal{L}(\tilde{s}) = w(\mathbf{S}^N) , \tag{2}$$

which relationship we will refer as *Blahut's Theorem*, since, as has been pointed out in [6], this relation was implicitly used by Blahut.

Example 3.1:
Let $F = GF(2)$ and let $\tilde{s} = s_0, s_1, s_2, \cdots$ be the characteristic phase of the PN-sequence defined by the primitive element α of $GF(2^3)$ whose minimum polynomial is $h(X) = X^3 + X + 1$, i.e., $\tilde{s}$ is defined by

$$s_i = Tr(\alpha^i) , \quad i \geq 0$$

where Tr is the trace operator from $GF(2^3)$ to $GF(2)$, i.e., $Tr(\alpha) = \alpha + \alpha^2 + \alpha^4$. Then $\tilde{s} = 1, 0, 0, 1, 0, 1, 1, 1, 0, 0, \cdots$. The sequence $\tilde{s}$ is periodic of period $N = 7$ and is generated by the LFSR with connection polynomial $C(D) = D^3 + D^2 + 1$ and initial state $[1, 0, 0]$. Consider now the time-domain vector $\mathbf{s}^N = [1, 0, 0, 1, 0, 1, 1]$ and its associated frequency-domain vector $\mathbf{S}^N = [S_0, S_1, \cdots, S_{N-1}]$ for the DFT defined by the primitive element $\omega = \alpha^{-1}$. Then $S_j = s(\omega^j)$, for $0 \leq j < N$, where $s(X) = 1 + X^3 + X^5 + X^6$ is the time-domain polynomial. Hence $\mathbf{S}^N = [0, 1, 1, 0, 1, 0, 0]$ and the frequency-domain polynomial is $S(X) = X + X^2 + X^4$, which is a so-called linearized polynomial [4].

4 PN-Sequences and their linearized polynomials

Suppose that α is a primitive element of $GF(2^L)$ whose minimum polynomial is $h(X) = X^L + c_1 X^{L-1} + \cdots + c_L$. Then the characteristic phase of the PN-sequence defined by $h(X)$ is the sequence $\tilde{s}$ such that

$$s_i = Tr(\alpha^i) = \alpha + \alpha^2 + \cdots + \alpha^{2^{L-1}} , \quad i \geq 0 . \tag{3}$$

The sequence $\tilde{s}$ is generated by the LFSR with connection polynomial $C(D) = 1 + c_1 D + \cdots + c_L D^L$ and initial state $[s_0, s_1, \cdots, s_{L-1}]$. Observe that $C(D)$ is the reciprocal polynomial of $h(X)$, so $\omega = \alpha^{-1}$ is (primitive and) a root of $C(D)$.
Thus

$$C(D) = \prod_{i \in \mathcal{C}} (1 - \alpha^i D)$$

where $\mathcal{C} = \{1, 2, 4, \cdots, 2^{L-1}\}$ is the main cyclotomic coset modulo $2^L - 1$. It follows from (3) and the inverse DFT relation

$$s_i = (\frac{1}{N^*}) S(\omega^{-i}) = S(\alpha^i)$$

that

$$S(X) = X + X^2 + X^4 + \cdots + X^{2^{L-1}} .$$

5 Second Order Nonlinear Filterings of PN-sequences

Let $\tilde{s} = s_0, s_1, s_2, \cdots$ be the characteristic phase of a PN-sequence generated by a maximal-length LFSR of length L and let $N = 2^L - 1$. Let $\tilde{s}_i$ denote the shifted version $s_{i-1}, s_i, s_{i+1}, \cdots$ for $i = 1, \cdots, N$ (in particular $\tilde{s} = \tilde{s}_1$). Note that, for $1 \leq i \leq L$, $\tilde{s}_i$ is the sequence that one would observe at the i-th stage of the LFSR when the LFSR was clocked repeatedly. Let $\mathbf{s}_i \in GF(2)^N$ be the binary vector corresponding to the first period of $\tilde{s}_i$. Let $w_2(i)$ denote the Hamming weight of the radix-2 form of the integer i.

We are interested in the linear complexity of sequences $\tilde{z}$ obtained by the product of the sequences produced from two different initial states of the maximal LFSR. Let $\tilde{t} = \tilde{s}_i \tilde{s}_j$ be the Hadamard product of $\tilde{s}_i$ and $\tilde{s}_j, i \neq j$, i.e., the bit-by-bit product of these two sequences. As was shown in section 4, defining the DFT by $\omega = \alpha^{-1}$ associates the frequency-domain polynomial $S(X) = X + X^2 + X^4 + \cdots + X^{2^{L-1}}$ with $\tilde{s}_1$. The shifting property 3.2.1 applied to $\tilde{s}_k$, $k \geq 1$, shows that the frequency-domain polynomial associated with $\tilde{s}_k$, $k \geq 1$, is $S(\alpha^{k-1} X)$, where α is the primitive element of $GF(2^L)$ used

to define the characteristic phase $\tilde{s}$ of the PN-sequence and $\omega = \alpha^{-1}$ is used to define DFT.

We use the notation $\tilde{s}_k \leftrightarrow S(\omega^{k-1}X)$ for the association given by the DFT. Then, for two distinct phases (say, $j > i$) $\tilde{s}_i$ and $\tilde{s}_j$ of $\tilde{s}$ we have $\tilde{s}_i \leftrightarrow S(\alpha^{i-1}X)$ and $\tilde{s}_j \leftrightarrow S(\alpha^{j-1}X)$. The convolution property (3.2.2), because $N^* = 1$, gives

$$\tilde{t} = \tilde{s}_i\tilde{s}_j \leftrightarrow T(X) = S(\alpha^{i-1}X)S(\alpha^{j-1}X) \ \ mod\ X^N - 1\ .$$

The linearized polynomial $S(\alpha^{i-1}X)$ has non-zero coefficients precisely for those powers e of X such that $w_2(e) = 1$ and $1 \le e < 2^L$. The product $S(\alpha^{i-1}X)S(\alpha^{j-1}X)$ thus has its potentially non-zero coefficients for those powers e of X such that $w_2(e) = 1$ or $w_2(e) = 2$ and $1 \le e < 2^L$. By Blahut's Theorem, the linear complexity of the sequence $\tilde{z}$ is just the number of non-zero coefficients of $T(X)$. Because $N = 2^L - 1$, all elements of a cyclotomic coset modulo N have the same Hamming weight since their L-bit radix-2 forms differ by just a cyclic shift. Thus

$$T(X) = \sum_{e \in I_2} T_e X^e$$

where I_2 is the union of the cyclotomic cosets $\mathcal{C}_s$ with $1 \le w_2(s) \le 2$. [For example, for $N = 7$, $\mathcal{C}_1 = \{1, 2, 4\}$ and $\mathcal{C}_3 = \{3, 6, 5\}$ are the only cyclotomic cosets mod N with weights 1 and 2.] Thus, because there are $\binom{L}{j}$ integers i, $1 \le i < 2^L$, such that $w_2(i) = j$, we have immediately the bound

$$\mathcal{L}(\tilde{s}_i\tilde{s}_j) \le L + \binom{L}{2}\ ,$$

which is Key's bound for second order filtering [3].

To find $\mathcal{L}(\tilde{t})$ exactly, it follows from the conjugacy constraints property (3.2.3) that it is enough to compute the coefficient T_e for those e that are representatives of the cyclotomic cosets modulo N. We now examine the coefficients T_e where $e = 1$ and $e = 2^{e_1} + 1$, $1 \le e_1 < L$, which are the representatives of the cyclotomic coset of weight 1 and those of weight 2. If we make the change of variables $Y = \alpha^{i-1}X$, then

$$T(\alpha^{-i+1}Y) = S(Y)S(\alpha^{j-i}Y) \ \ mod\ Y^N - 1\ .$$

Define the coefficients A_i, $i \in \mathcal{C}$, by

$$S(\alpha^{j-i}Y) = \sum_{i \in \mathcal{C}} A_i Y^i$$

where $\mathcal{C}$ is the main coset modulo N, i.e., $\mathcal{C} = \{1, 2, 4, \cdots, 2^{L-1}\}$.

The coefficient of Y in $S(Y)S(\alpha^{j-i}Y) \ \ mod\ Y^N - 1$ and in $T(\alpha^{-i+1}Y)$ is

$$A_{2^{L-1}} = (A_1)^{2^{L-1}} = (\alpha^{j-i})^{2^{L-1}} \neq 0 \ \text{ and } \ T_1\alpha^{-i+1},$$

respectively. This implies that T_1 is non zero. For the cosets of weight 2, we consider the coefficient of $Y^{2^{e_1}+1}$ in $S(Y)S(\alpha^{j-i}Y)$ *mod* $Y^N - 1$ and in $T(\alpha^{-i+1}Y)$, which is

$$A_1 + A_1^{2^{e_1}} \text{ and } \alpha^{-i+1}T_{2^{e_1}+1} .$$

We conclude that $T_{2^{e_1}+1}$ will be zero if and only if $A_1 = \alpha^{j-i} \neq 1$ is a zero of the polynomial $X^{2^{e_1}} + X$, which happens if and only if e_1 divides L and $e_1 > 1$

We have proved the following result:

Proposition 1 *Let $\tilde{s}_i$ and $\tilde{s}_j$, $1 \leq i < j < 2^L - 1$ be distinct phases of the same sequence $\tilde{s}$ whose minimal polynomial $h(X) \in GF(2)[X]$ is primitive and has degree L, L a prime. Let $\alpha \in GF(2^L)$ denote a root of $h(X)$, and let $\tilde{t} = \tilde{s}_i\tilde{s}_j$ be the Hadamard product of the two distinct phases. Then*

$$\mathcal{L}(\tilde{t}) = L + \binom{L}{2} .$$

The above arguments show in general that any sum of Hadamard products of order n of the sequences $\tilde{s}_i$, $i = 1, 2, \cdots, N$, is a sequence $\tilde{t}$ for which $T(X)$ has $\binom{L}{n}$ potentially non-zero coefficients. Thus $\tilde{t}$ has linear complexity at most $\binom{L}{n}$. The linear complexity of a sequence $\tilde{t}$ obtained by k-th order nonlinear filtering of $\tilde{s}$ then satisfies

$$\mathcal{L}(\tilde{t}) = \sum_{n=1}^{k} \binom{L}{n} ,$$

which is Key's general bound.

6 Conclusion

We have given a method for analyzing the linear complexity of nonlinear filterings of PN-sequences that is based on the Discrete Fourier Transform.
As an application to show the usefulness of our approach, we gave a simple proof of Key's upper bound on linear complexity of nonlinearly filtered PN-sequences and we further showed that Key's result, viz. that products of two distinct phases of the same PN-sequence of period $2^L - 1$ always have the maximum possible linear complexity $\binom{L}{1} + \binom{L}{2}$ for such products when L is prime; the restrictions on phase differences imposed by this author for this nondegeneracy of linear complexity are unnecessary.

References

[1] R. E. Blahut, *Theory and Practice of Error Control Codes*, Addison-Wesley Publishing Company, 1983.

[2] E. J. Groth "Generation of binary sequences with controllable complexity," *IEEE Trans. Inform. Theory*, vol. IT-17, no. 3, pp.288-296, May 1971.

[3] E. L. Key "An analysis of the structure and complexity of nonlinear binary sequence generators," *IEEE Trans. Inform. Theory*, vol. IT-22, no. 6, pp.732-736, Nov. 1976.

[4] R. Lidl and H. Niederreiter, *Introduction to Finite Fields and their Applications*, Cambridge: Cambridge Univ. Press, 1986.

[5] J. L. Massey, "Shift-register synthesis and BCH decoding," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 122-127, Jan. 1969.

[6] J. L. Massey, "Review of R. E. Blahut, *Theory and Practice of Error Control Codes*", *IEEE Trans. Inform. Theory*, vol. IT-31, no. 4, p. 553, Jul. 1985.

[7] R. A. Rueppel, *Analysis and Design of Stream Ciphers*, Berlin: Springer-Verlag, 1986.