Signal Processing Techniques in Cryptography

Konstantinos Limniotis*

Department of Informatics and Telecommunications
National and Kapodistrian University of Athens
TYPA Buildings, University Campus, 15784 Athens, Greece
klimn@di.uoa.gr

Abstract. Security of cryptographic symmetric primitives is studied in this thesis. Pseudorandomness characteristics of cryptographic sequences are analyzed, resulting in new methods for constructing sequences with high linear complexity. Connections between nonlinear complexity and other cryptographic criteria are also established, whereas a new recursive algorithm for efficiently computing the minimal feedback shift register which generates a given sequence is provided. Furthermore, security issues of cryptographic Boolean functions that are used in cryptographic systems as components of sequence generators are studied; on this direction, new efficient formulas for determining best quadratic approximations of several classes of Boolean functions are derived, leading to new design principles that should be considered in the construction of secure cryptosystems.

Keywords: Boolean functions, complexity, Discrete Fourier Transform, feedback shift registers, sequences, stream ciphers.

1 Introduction

Cryptographic algorithms are categorized into two families, namely symmetric or secret-key algorithms and public-key algorithms [22]. Symmetric algorithms are the only ones that achieve several important functionalities such as high speed and low-cost encryption and are used in conjunction with public-key techniques in order to safely distribute the secret key among the members of a group. Symmetric algorithms, being further classified into block ciphers and stream ciphers, are used in many applications. Especially stream ciphers are widely used to provide confidentiality in environments characterized by a limited computing power or memory capacity, and the need to encrypt at high speed. Typical examples of stream ciphers are the A5/1 and E0 algorithms, employed in GSM communications and Bluetooth protocol respectively.

In general, a stream cipher consists of a binary keystream generator, whose output $k_1k_2\ldots$ is added modulo 2 to the original message $m_1m_2\ldots$, leading to the encrypted message (ciphertext) $c_1c_2\ldots$ (Fig. 1). *Shift registers* of linear (LFSR) or nonlinear (NFSR) feedback are a basic building block of keystreams generators in stream ciphers. The security of such systems is strongly contingent on the pseudorandomness characteristics of the keystream. The pseudorandomness is attributed to several factors; amongst others, an important cryptographic

* Dissertation Advisor: Nicholas Kalouptsidis, Professor.

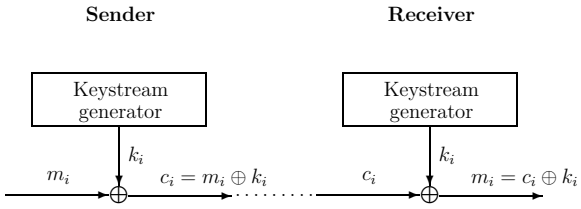

Fig. 1. Basic functionality of a stream cipher

feature of a sequence is its *nonlinear complexity* or simply *complexity*, defined as the length of the shortest feedback shift register that produces the sequence. Especially the *linear complexity*, i.e. the length of the shortest LFSR generating a given sequence, is important for assessing resistance to cryptanalytic attacks, like the *Berlekamp–Massey algorithm* (BMA) [19]. However, determining the connections and trade-offs between several cryptographic criteria of sequences remains an open problem. Since most constructions are ad-hoc, finding good generators is of great theoretical and practical value.

High linear complexity keystreams are generated by applying Boolean functions either as *filters* [7][26] or *combiners* [26], to one or several LFSRs respectively. In any case, the highest value attainable by linear complexity depends on the degree of the function [7]. The problem of determining the exact linear complexity attained by filterings is open. Two classes of filters have been introduced, namely *equidistant* [26] and *normal* [8], that allow to derive lower bounds on the linear complexity. Given a filter of degree k and a LFSR of length n whose characteristic polynomial is primitive over $\mathbb{F}_{2^n}$, the lower bound on the linear complexity of keystreams is $\binom{n}{k}$ for both types of filters. These results rely on the so-called Rueppel's *root presence test* [26].

On the contrary, the general case of nonlinear complexity has not been studied to the same extent. In [4], a directed acyclic word graph is used to exhibit the complexity profile of sequences over arbitrary fields. An approximate probability distribution for the nonlinear complexity of random binary sequences is derived in [3]. Recent results are provided in [24], where the minimal nonlinear FSR generating a given sequence is computed via an algorithmic approach, and [25] where the special case of a quadratic feedback function of the FSR is treated.

Apart from the pseudorandomness characteristics of keystream, many attacks on conventional cryptographic algorithms are related to some properties of the underlying Boolean functions. The formalization of well-known attacks against LFSR-based stream ciphers have led to the definitions of some relevant quantities related to Boolean functions. These quantities measure the resistance of a cryptosystem to classical attacks. For instance, high algebraic degree of nonlinear filters or combiners is prerequisite for constructing sequences achieving high linear complexity. Furthermore, the *nonlinearity* of Boolean functions is one of the most significant cryptographic properties; it is defined as the minimum distance from all affine functions, and indicates the degree to which attacks based on linear cryptanalysis [21] can be prevented. With the appearance of more recent attacks, such as algebraic [2], and low order approximation attacks [11],

Boolean functions need also have the property that they cannot be approximated efficiently by low degree functions. Hence, the *rth order nonlinearity* characteristics of Boolean functions need also be analyzed. This is known to be a difficult task for $r > 1$, whereas even the second order nonlinearity is unknown for all Boolean functions, with the exception of some special cases, or if the number of variables n is small [1].

In this thesis, state-space representations are employed as vehicle to the study of complexity of binary sequences. System theoretic concepts, namely controllability and observability, are used to characterize minimal sequence generators [6]. Jordan canonical forms are used for the complete analysis of sequences whose Fourier transform is not defined [15]. A new *generalized discrete Fourier transform* (GDFT) is proposed that presents the same properties with the GDFT defined in [20]. In addition, connections of this new GDFT with a new vectorial trace representation of sequences are established that *facilitate the generation of sequences with prescribed linear complexity*. Furthermore, nonlinearly filtered maximal length sequences with period $N = 2^n - 1$ are studied under this framework, resulting in new general classes of nonlinear filters of degree k which *generalize Rueppel's equidistant filters and guarantee the same lower bound* $\binom{n}{k}$ *on the linear complexity* [15]. The connections between the nonlinear and Lempel-Ziv complexity are also studied, which is a well-known open problem [23]. It is shown that the eigenvalue profile of a sequence, which determines the Lempel-Ziv complexity, also determines its nonlinear complexity profile [14]. Furthermore, for any periodic binary sequence, we establish the dependence of the minimum achievable compression ratio on its nonlinear complexity by deriving a lower bound depending on the complexity [14]. Based on the properties of the nonlinear complexity profile, a new efficient recursive algorithm producing the minimal FSR of a binary sequence is developed, thus *generalizing the BMA to the nonlinear case* [14],[16]. Finally, explicit formulas are proved that compute all best quadratic approximations of a class of functions with degree 3 or 4 [9]. These results are based upon Shannon's expansion formula of Boolean functions and *hold for an arbitrary number n of variables*. The derived method reveals new design principles for cryptographic functions. An analysis of contemporary constructions of functions is also performed, indicating potential weaknesses if construction parameters are not properly chosen.

This summary is organized as follows; First, Section 2 introduces the basic definitions and settles the notation. Section 3 provides the basic results regarding the linear complexity of sequences obtained by state space generators, while Section 4 presents the new results regarding the nonlinear complexity, as well as its connections with Lempel-Ziv complexity. The algorithmic method of computating the best quadratic approximations of Boolean functions is described in Section 5. Finally, concluding results are given in Section 6.

2 Preliminaries

Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be a Boolean function, where $\mathbb{F}_2 = \{0, 1\}$. The set of Boolean functions on n variables is denoted by $\mathbb{B}_n$. The complement of a binary variable x will be denoted by $x' = x + 1$, where "+" represents addition modulo 2.

Boolean functions are expressed in their *algebraic normal form* (ANF) as

$$f(x_1, \ldots x_n) = \sum_{e \in \mathbb{F}_2^n} a_e \, x_1^{e_1} \cdots x_n^{e_n}, \qquad a_e \in \mathbb{F}_2 \tag{1}$$

where the sum is taken modulo 2, $e = (e_1, \ldots, e_n)$, while $x_i^1 = x_i$ and $x_i^0 = 1$. The *degree* of f equals $\deg(f) = \max\{\mathrm{wt}(e) : a_e = 1\}$, and $\mathrm{wt}(e)$ is the *Hamming weight* of vector e. If $\deg(f) = 1, 2, 3$, then f is called *affine* (or *linear* if its constant term is zero), *quadratic*, *cubic*; terms with degree $k \leq \deg(f)$ in ANF comprise its *kth degree part*. The *distance* of $f, g \in \mathbb{B}_n$ is $\mathrm{wt}(f + g)$.

Another representation , the so-called *Exclusive-or Sum-Of-Products* (ESOP), occurs if the variables in (1) are taken to be in either complemented or uncomplemented form.

The *Shannon's expansion formula* of $f \in \mathbb{B}_n$ with respect to x_j is

$$f(x_1, \ldots, x_n) = f_0 \parallel_j f_1 \triangleq (1 + x_j) f_0 + x_j f_1 \,, \quad 1 \leq j \leq n$$

where *sub-functions* $f_0, f_1 \in \mathbb{B}_{n-1}$ do not depend on x_j; they are the restriction of f in $x_j = 0, 1$.

The Walsh transform of $f \in \mathbb{B}_n$ at $a \in \mathbb{F}_2^n$ is the real-valued function

$$\widehat{\chi}_f(a) = \sum_{x \, \in \, \mathbb{F}_2^n} \chi_f(x) (-1)^{\phi_a(x)} = 2^n - 2 \, \mathrm{wt}(f + \phi_a) \tag{2}$$

with $\chi_f(x) = (-1)^{f(x)}$ and $\phi_a(x) = \sum_i a_i x_i$. The minimum distance between f and all affine functions is determined by

$$\mathcal{NL}_f = \min_{v \, \in \, \mathfrak{R}(1,n)} \big\{ \mathrm{wt}(f + v) \big\} = 2^{n-1} - \frac{1}{2} \max_{a \, \in \, \mathbb{F}_2^n} |\widehat{\chi}_f(a)| \tag{3}$$

and is called *nonlinearity* of f. An affine function v such that $\mathrm{wt}(f + v) = \mathcal{NL}_f$ is a *best affine approximation* of f, denoted by λ_f, and $\mathcal{A}_f$ is the set of all its best affine approximations. The above can be extended to *best quadratic approximations* of f, denoted by ξ_f, which are quadratic functions u satisfying $\mathrm{wt}(f + u) = \min_{u : \deg(u) \leq 2} \{\mathrm{wt}(f + u)\} \triangleq \mathcal{NQ}_f$.

A sequence $y = \{y_i\}_{i \geq 0}$ with elements in the finite field $\mathbb{F}_2$ is said to be *ultimately periodic* if there exist integers $T > 0$ and $t_0 \geq 0$ such that $y_{i+T} = y_i$ for all $i \geq t_0$. The least integer T with this property is called *period* of y, and t_0 is its *preperiod*. If $t_0 = 0$, then the sequence y is said to be *periodic*. If y has finite length N, then $y^N \triangleq y_0^{N-1}$ denotes the whole sequence. For any $0 \leq j < N$, the tuple y_0^j is a *prefix* of y^N; for the special case that $j < N - 1$, such a prefix is called *proper prefix*. A *suffix* of y^N is any tuple y_j^{N-1}, $0 \leq j \leq N - 1$; a proper suffix is similarly defined. Such sequences are typically generated by FSRs satisfying a recurring relation of the form $y_{i+n} = h(y_{i+n-1}, \ldots, y_i)$, $i \geq 0$, where $n > 0$ equals the number of stages of the FSR. The feedback $h : \mathbb{F}_2^n \to \mathbb{F}_2$ is a nonlinear function, usually having a zero constant term, mapping elements of the nth-dimensional vector space $\mathbb{F}_2^n$ onto $\mathbb{F}_2$. In the case of a linear feedback, i.e. $y_{i+n} = a_{n-1} y_{i+n-1} + \cdots + a_1 y_{i+1} + a_0 y_i$, each LFSR is associated with its characteristic polynomial $f(z) = z^n + a_{n-1} z^{n-1} + \cdots + a_0$.

Definition 1 *The length of the shortest FSR generating a sequence y^N is referred to as* nonlinear complexity *of y^N, and is denoted by $c(y^N)$. The integer-valued sequence $c(y^1), \ldots, c(y^N)$ is called* nonlinear complexity profile. *The linear complexity* $\mathrm{lc}(y^N)$ *and the linear complexity profile are similarly defined.*

When the LFSR generating y has the minimum possible length, then its characteristic polynomial $f(z)$ is called *minimal polynomial* of y.

Let the vector $\boldsymbol{y} = (y_0\ y_1\ \cdots\ y_{N-1})$ contain the first N elements of a binary periodic sequence y of period N. Let N be a divisor of $2^n - 1$ for some positive integer n. The Fourier transform of $\boldsymbol{y}$ is the vector $\boldsymbol{Y} = (Y_0\ Y_1\ \cdots\ Y_{N-1})$ of length N whose elements are given by

$$Y_j = \sum_{i=0}^{N-1} y_i \alpha^{ij}, \qquad 0 \le j < N \tag{4}$$

where $Y_j \in \mathbb{F}_{2^n}$ and $\alpha \in \mathbb{F}_{2^n}$ is an element of order N in the extension field $\mathbb{F}_{2^n}$. Vector $\boldsymbol{y}$ is reconstructed from $\boldsymbol{Y}$ by means of the inverse Fourier transform

$$y_i = \sum_{j=0}^{N-1} Y_j \alpha^{-ij}, \qquad 0 \le i < N\,. \tag{5}$$

A direct consequence of (4) is that the Fourier coefficients satisfy the *conjugacy property*, i.e. $Y_{2j \bmod N} = Y_j^2$ for all $0 \le j < N$. The linear complexity of a periodic sequence over $\mathbb{F}_2$, with period N a divisor of $2^n - 1$ for some integer n, equals the Hamming weight of its Fourier transform (*Blahut's theorem*). A generalized Fourier transform, that describes sequences of arbitrary period, is defined in [20].

For any positive integer N such that $\gcd(N, 2) = 1$ and each $j \in \mathbb{Z}_N = \{0, \ldots, N-1\}$, we define the set of distinct elements $I_j = \{j, 2j, \ldots, 2^{n_j - 1} j\}$ to be the *cyclotomic coset* of j, where all elements are taken modulo N and $n_j = |I_j|$. The least element in I_j is referred to as *coset leader* and the set containing all coset leaders modulo N will be denoted by I. From the definition of cyclotomic cosets and the conjugacy property satisfied by (4) and (5), the Fourier transform can be equivalently written as

$$y_i = \sum_{j \in I} \mathrm{tr}_1^{n_j}(Y_j \alpha^{-ij}) \tag{6}$$

where $Y_j \in \mathbb{F}_{2^{n_j}}$ and the function $\mathrm{tr}_1^{n_j}(z) = z + z^2 + \cdots + z^{2^{n_j - 1}}$ is the *trace function* that maps elements of $\mathbb{F}_{2^{n_j}}$ onto its prime subfield $\mathbb{F}_2$ [13]. The above is called *trace representation* of the sequence y.

3 Linear complexity of sequences obtained by state-space generators

In this section we focus on linear state space generators, described by

$$x_{i+1} = \boldsymbol{A}\, x_i \tag{7a}$$

$$y_i = c^T\, x_i \tag{7b}$$

where x_i, c are $n \times 1$ vectors, c^T denotes the transpose of c, and $\boldsymbol{A}$ is an $n \times n$ matrix. The integer n defines the *dimension* of the system $\mathfrak{L}$. Clearly, any LFSR can be described by (7). Any such system is denoted by $\mathfrak{L} = \langle \boldsymbol{A}, c, x_0 \rangle$.

Proposition 1 ([5]) *A linear realization $\mathfrak{L} = \langle \boldsymbol{A}, c, x_0 \rangle$ of a periodic sequence y with dimension n is minimal (i.e. there is no other linear realization of lower dimension generating y) if and only if it is both controllable and observable.*

Proposition 2 ([5]) *Let $\mathfrak{L} = \langle \boldsymbol{A}, c, x_0 \rangle$ and $\mathfrak{L}' = \langle \boldsymbol{A}', c', x_0' \rangle$ be two minimal linear realizations of a periodic sequence y. Then, $\mathfrak{L}$ and $\mathfrak{L}'$ are necessarily isomorphic (or equivalent) since there exists a change of coordinates $\boldsymbol{P}$ such that it holds $\boldsymbol{A}' = \boldsymbol{P}\boldsymbol{A}\boldsymbol{P}^{-1}$, $(c')^T = c^T\boldsymbol{P}^{-1}$, and $x_0' = \boldsymbol{P}x_0$.*

It is well-known that any matrix with coefficients in an algebraically closed field can be put into the so-called *Jordan canonical form*. Thus, among all isomorphic minimal linear realizations $\mathfrak{L}$ of a given sequence y, there exists one with state transition matrix $\boldsymbol{A}$ in the Jordan canonical form.

Theorem 1 ([15]) *Let $\mathfrak{L} = \langle \boldsymbol{A}, c, x_0 \rangle$ be a linear realization of sequence y with matrix $\boldsymbol{A}$ in the Jordan canonical form. The generator $\mathfrak{L}$ is controllable (resp. observable) if and only if there is one Jordan block associated with each eigenvalue and all elements of the initial state vector x_0 (resp. output vector c) corresponding to the last row (resp. first column) of each Jordan block are nonzero.*

Theorem 2 ([15]) *With the above notation, let the state transition matrix $\boldsymbol{A}$ be diagonal. Then, the generator $\mathfrak{L}$ is minimal if and only if the eigenvalues of $\boldsymbol{A}$ are pairwise distinct and all elements of x_0 and c are nonzero.*

In this thesis it is proved (by using the above results) that, if y is a periodic binary sequence with least period N, then it admits a diagonal realization $\mathfrak{L}$ over the splitting field of $z^N - 1$ if and only if its Fourier transform exists, or equivalently $\gcd(N, 2) = 1$. In this case, the initial state of the diagonal realization with dimension N equals its Fourier transform. Hence, we directly prove that the dimension of a minimal realization of any such sequence equals its linear complexity. Similar results for the more interesting general case of sequences whose Fourier transform is not defined are also proved. Let $z = (z_1 \ z_2 \ \cdots \ z_m)^T$ be an $m \times 1$ vector with elements over $\mathbb{F}_{2^n}$. We define the *block-trace function*

$$\mathbf{tr}_1^n(z) = \left(\mathrm{tr}_1^n(z_1) \ \mathrm{tr}_1^n(z_2) \ \cdots \ \mathrm{tr}_1^n(z_m)\right)^T \tag{8}$$

that maps vectors of $\mathbb{F}_{2^n}^m$ to elements of the vector space $\mathbb{F}_2^m$. Then, we prove the following.

Theorem 3 *Let y be a binary sequence of period $N = 2^e m$, with m odd and $e > 0$, and minimal polynomial $f(z)$ factored as*

$$f(z) = f_1(z)^{d_1} f_2(z)^{d_2} \cdots f_r(z)^{d_r} \tag{9}$$

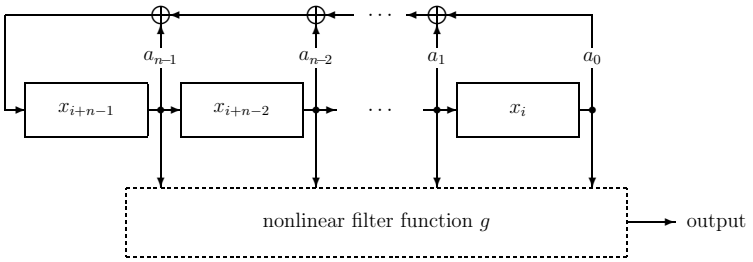

Fig. 2. A nonlinear filter applying to a LFSR

where $d_s > 0$ and f_s is irreducible of degree n_s, $1 \leq s \leq r$. Further let α be a primitive mth root of unity over $\mathbb{F}_2$ lying in the splitting field of $z^N - 1$, and let α^{j_s} be a root of f_s. Then y is always written in the vectorial trace representation

$$y_i = \sum_{1 \leq s \leq r} \mathbf{1}_{d_s}^T \, \mathbf{tr}_1^{n_s}(\boldsymbol{J}_{j_s}^i z_s), \qquad i \geq 0 \tag{10}$$

where $z_s = (z_{s,1} \; z_{s,2} \; \cdots \; z_{s,d_s})$ is a vector over the splitting field of $f(z)$, and the Jordan block $\boldsymbol{J}_{j_s}$ has dimension d_s and diagonal element α^{j_s}.

Equation (10) provides a novel generalized trace representation of any sequence y, including those whose minimal polynomial does not have distinct irreducible factors. The practical importance of the above resides in the fact that we can generate sequences of prescribed linear complexity by appropriately selecting the initial state *for any LFSR*.

Based on (10), which generalizes (6), we introduce below a new *generalised discrete Fourier transform* (GDFT) for sequences of arbitrary period $N = 2^e m$, with m odd and $e > 0$.

Definition 2 ([15]) *Let y be a binary sequence of period $N = 2^e m$ generated by $\mathfrak{L} = \langle \boldsymbol{J}, \mathbf{1}, \boldsymbol{Y} \rangle$ of dimension N, where $\boldsymbol{J}$ is a Jordan matrix. Then, the initial state*

$$\boldsymbol{Y} = (Y_0 \; Y_1 \; \cdots \; Y_{N-1})^T \tag{11}$$

is defined as the generalized discrete Fourier transform of y.

The above definition is easily generalized to sequences over fields with an odd prime characteristic p, since it can be easily seen that the vectorial trace representation (10) holds for sequences over any field. Clearly, if $e = 0$ the above generalized discrete Fourier transform and the vectorial trace representation (10) coincide with their ordinary counterparts. The advantage of the proposed GDFT is that while being a natural generalization of the usual DFT from a system theoretic point of view, it also allows the easy computation of the linear complexity of sequences by means of the Günther weight, like the GDFT proposed in [20].

Next, we consider a LFSR of length n with a primitive characteristic polynomial f, where a nonlinear filter function g of degree $k \leq n$ is applied to its stages

(Fig. 2). The realization $\mathfrak{N} = \langle \boldsymbol{A}, g, x_0 \rangle$ of the sequence $y = \{y_i\}_{i \geq 0}$ generated by the above automaton is described by

$$x_{i+1} = \boldsymbol{A} x_i \tag{12a}$$

$$y_i = g(x_i) \tag{12b}$$

where $\boldsymbol{A}$ is the companion matrix of the characteristic polynomial f and $x_0 = (x_{0,1}\ x_{0,2}\ \cdots\ x_{0,n})^T$ is the initial state of $\mathfrak{N}$. The study of such generators is simplified if they are linearly described; this is accomplished by treating each product term in the ANF of g as a single variable by suitably extending the state space. This procedure is called *linearization* of generator $\mathfrak{N}$ and is based upon properties of Kronecker products. Hence, minimality characteristics of the equivalent linearized system are also determined by controllability and observability arguments. It is proved in this thesis that the Rueppel's root presence test can be re-derived by this analysis. Furthermore, based on the analysis via Kronecker products, the following result is proved.

Theorem 4 ([15]) *Let $\mathfrak{N} = \langle \boldsymbol{A}, g, x_0 \rangle$ be a generator of dimension $n > 0$ and let the characteristic polynomial f of the state transition matrix $\boldsymbol{A}$ be primitive. For some positive integer δ with $\gcd(\delta, 2^n - 1) = 1$ assume the nonlinear filter g consists of only one product of degree $k < n$*

$$g(z_1, z_2, \ldots, z_n) = z_{t_1} z_{t_2} \cdots z_{t_k}$$

where $t_1 > \delta$ or $t_k \leq n - \delta$. If there exists an integer $1 \leq i \leq k$ such that the product $g_i(z_1, \ldots, z_n) = z_{t_1} \cdots z_{t_{i-1}} z_{t_{i+1}} \cdots z_{t_k}$ is equidistant, with distance δ, then the linear complexity of the generated sequence is lower bounded by $\binom{n}{k}$.

Theorem 4 generalizes Rueppel's results since it defines filter functions that are slightly different from equidistant filters but admit the same lower bound on the linear complexity. It is also proved that these results can be used to generalize other classes of nonlinear filters that generate sequences achieving a prescribed lower bound on the linear complexity - such as filters based on normal bases. Clearly, the above new construction provides greater flexibility in designing nonlinear filters that output sequences of high linear complexity.

4 Nonlinear complexity and Lempel-Ziv complexity

This section studies the nonlinear conplexity of sequences and its connections with Lempel-Ziv complexity; the latter is defined by the number of words occuring via a specific parsing procedure of the sequence described in [12]. As it is shown in [12], the Lempel-Ziv complexity is determined by the the eigenvalue profile $k(y^1), k(y^2), \ldots, k(y^N)$ of the sequence y^N. The value of $k(y^i)$ equals $i - s_i$, where s_i is the length of the longest suffix of y^i that is present at least twice within y^i. By proving a series of results, we get the following.

Theorem 5 ([14]) *Let $c(y^{n-1}) = m$ and assume the minimal FSR of y^{n-1} does not generate y^n. Then, it holds*

$$c(y^n) = \max\{c(y^{n-1}), n - k(y^{n-1})\} \tag{13}$$

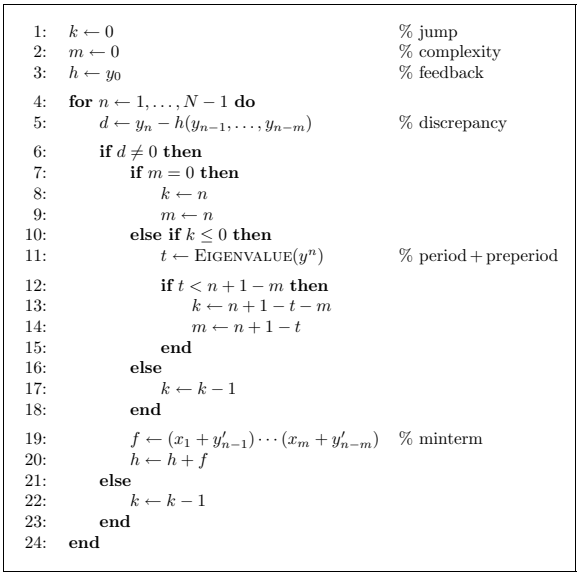

Fig. 3. A recursive algorithm for minimal FSR synthesis of binary sequence y^N

The next result exhibits that the eigenvalue profile uniquely determines the complexity profile.

Theorem 6 ([14]) *If two sequences have the same eigenvalue profile, then they necessarily have the same nonlinear complexity profile.*

The above are used to develop a recursive algorithm that computes the minimal FSR of any binary sequence, which comprises the generalization of the BMA to the nonlinear case. This algorithm is illustrated in Fig. 3 [14],[16]. The feedback function of the minimal FSR is given in the ESOP representation. In Lines 11, 12 of the algorithm, we examine whether a jump in the complexity occurs based on Theorem 5; if a jump occurs, then its value is computed in Line 13. Note that this step has linear computational complexity due to the existence of the *Knuth-Morris-Pratt* (KMP) algorithm for pattern matching. The computational complexity of the algorithm mainly rests with line 5 where the boolean function $h^{(n)}$ is evaluated. For each $n \leq N$ the ESOP representation of $h^{(n)}$ has less than n terms, each consisting of at most $c(y^n) \leq n$ variables. Since no term is present in the ESOP representation of $h^{(n)}$ having more than $c(y^N)$ variables, the computational complexity of the algorithm in the average case highly depends on the expected value of the nonlinear complexity of random binary sequences of given length N. It is known that for large N it holds $E(c(y^N)) \approx 2\log_2 N$, and the average computational complexity of the algorithm becomes $O(N^2 \log_2 N)$. Note that our algorithm has the same computational complexity with the one proposed in [4] for determining the minimal FSR of a given sequence. However, its recursive nature is an important advantage since it eliminates the need to know the entire sequence in advance.

Connections between nonlinear complexity and Lempel-Ziv compression ratio are also established. More precisely, the next result is proved.

Theorem 7 ([14]) *Let y be a binary sequence with period N, and let $c(y) = m$. If y^N denotes the first N terms of y and n is the largest integer such that $2^n \leq m < 2^{n+1}$, then*

$$\rho_{y^N} > \min\left\{ \frac{1}{m}\lceil \log_2(2m) \rceil, \frac{1}{2^n}(n+1) \right\}, \tag{14}$$

where ρ_{y^N} is the compression ratio of y^N according to the Lempel-Ziv compression algorithm of [27].

As it is proved in this thesis, the bound in (14) decreases as the complexity $c(y)$ of the periodic sequence y increases. Therefore, it is possible to design a construction for generating sequences of very high complexity, which however are highly compressible. Since truly random sequences do not present such behavior, we expect that compressibility is used in conjunction with nonlinear complexity to filter out sequences having this type of deficiency. A specific class of sequences achieving high nonlinear complexity and low compression ratio is identified in this thesis, the so-called s-optimal sequences, thus revealing the cryptographic value of compressibility.

5 Efficient computations of best quadratic approximations of Boolean functions

This section presents new efficient formulas for determining best quadratic approximations of boolean functions. The main result is the following.

Theorem 8 ([9]) *Let $f \in \mathbb{B}_n$ be a cubic function, where there exists variable x_j such that $f = (q + l_0) \parallel_j (q + q_j + l_1)$ (q, q_j are quadratic). Then, the best quadratic approximations of f have one of the following forms*

i. $\xi_f^0 = (q + l_0) \parallel_j (q + l_1 + \lambda_{q_j})$;
ii. $\xi_f^1 = (q + q_j + l_0 + \lambda_{q_j}) \parallel_j (q + q_j + l_1)$.

Corollary 1. *The second order nonlinearity of any cubic function $f \in \mathbb{B}_n$ of the above form is equal to $\mathcal{NQ}_f = 2^{n-2} - 2^{n-2-h_{q_j}}$, for some $1 \leq h_{q_j} \leq \lfloor (n-1)/2 \rfloor$.*

The above is proved by means of special properties that characterize the Walsh transform of quadratic boolean functions. The importance of Theorem 8 rests with the fact that it enables direct computation of all the best quadratic approximations of a particular subset of cubic Boolean functions on n variables, which have a variable being present in all cubic terms, by determining the best affine approximations of quadratic Boolean functions on $n-1$ variables; direct formulas for determining these best affine approximations are also proved in this thesis, which exploit the representation of quadratic functions according to Dickson's theorem [18], without using the Walsh transform. Cubic functions of the form described in Theorem 8 have been recently proposed for contemporary stream ciphers, thus revealing the cryptographic importance of the above result.

Best quadratic approximations of functions with degree 4 are also proved.

Theorem 9 ([9]) *Let $f \in \mathbb{B}_n$ be a Boolean function of degree 4, and let $f = f_0 \parallel_j f_1$, for some $1 \leq j \leq n$, such that f_0 is cubic function of the form described in Theorem 8 and $f_1 = q + l$ is a quadratic function, where q, l are its quadratic and linear part respectively. If $\mathcal{NL}_{f_0+q} \leq 2^{n-2} - 2^{n-4}$, then all functions*

$$g = (q + \lambda_{f_0+q}) \parallel_j f_1 \tag{15}$$

are best quadratic approximations of f and $\mathcal{NQ}_f = \mathcal{NL}_{f_0+q}$. Otherwise, it holds $\mathcal{NQ}_f > 2^{n-2} - 2^{n-4}$.

It is evident from the above that constructions of Boolean functions based on the concatenation of low-degree functions with fewer number of variables are susceptible to successful best quadratic approximation attacks if the sub-functions are not properly chosen, and in particular if the resulting Boolean function has low second order nonlinearity. Since many constructions of bent functions or correlation-immune functions (both admitting important cryptographic properties) present this structure, our results determine new design principles that need to be considered in constructions of boolean functions, so as to guarantee resistance in low order approximation attacks [10].

6 Conclusions

This thesis studies cryptographic features of sequences and Boolean functions, by using signal processing techniques. This leads to new methods for constructing cryptographic primitives achieving good cryptographic properties. Research is progress focuses on generalizing the results of Section 5 in a wider class of Boolean functions, while the security of several contemporary stream ciphers with respect to low order approximations is also currently studied. Moreover, the connection between several other cryptographic criteria of sequences, apart from nonlinear and Lempel-Ziv complexity, remains an interesting open problem.

References

1. Carlet, C.: Recursive lower bounds on the nonlinearity profile of Boolean functions and their applications. Cryptology ePrint Archive, Report 2006/459 (2006) http://eprint.iacr.org.
2. Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback. Advances in Cryptology - Eurocrypt '03 (Lecture Notes in Computer Science, Springer-Verlag) **2656** (2003) 345–359.
3. Erdmann, D., Murphy, S.: An approximate distribution for the maximum order complexity. Des. Codes and Cryptography **10** (1997) 325–339.
4. Jansen, C. J., Boekee, D. E.: The shortest feedback shift register that can generate a given sequence. Proc. Advances in Cryptology-CRYPTO '89 (1990) 90–99.
5. Kalouptsidis, N.: Signal Processing Systems. Telecommunications and Signal Processing Series, John Wiley & Sons (1996)
6. Kalouptsidis, N. and Limniotis, K.: Nonlinear span, minimal realizations of sequences over finite fields and De Brujin generators. Proc. Int. Symp. Inf. Theory and Appl., (2004) 794–799.

7. Key, E. L.: An analysis of the structure and complexity of nonlinear binary sequence generators. IEEE Trans. Inform. Theory **22** (1976) 732–736.
8. Kolokotronis, N., Kalouptsidis, N.: On the linear complexity of nonlinearly filtered PN-sequences. IEEE Trans. Inform. Theory **49** (2003) 3047–3059.
9. Kolokotronis, N., Limniotis, K. and Kalouptsidis, N.: Best affine approximations of boolean functions and applications to low order approximations. Proc. IEEE Int. Symp. Inf. Theory (2007) 1836–1840.
10. Kolokotronis, N., Limniotis, K. and Kalouptsidis, N.: Efficient computation of the best quadratic approximations of cubic boolean functions. 11th IMA International Conference on Cryptography and Coding, (Lecture Notes in Computer Science, Springer-Verlag) **4887** (2007) 73–91.
11. Kurosawa, K., Iwata, T., Yoshiwara, T.: New covering radius of Reed-Muller codes for t-resilient functions. IEEE Transactions on Information Theory **50** (2004) 468–475.
12. Lempel, A., Ziv, J.: On the complexity of finite sequences. IEEE Trans. Inform. Theory **22** (1976) 75–81.
13. Lidl, R., Niederreiter, H.: Finite Fields. vol. 20 of Encyclopedia of Mathematics and its Applications. Cambridge University Press 2nd ed. (1996)
14. Limniotis, K., Kolokotronis, N., and Kalouptsidis, N.: On the nonlinear complexity and Lempel–Ziv complexity of finite length sequences. IEEE Trans. Inform. Theory **53** (2007) 4293–4302.
15. Limniotis, K., Kolokotronis, N., and Kalouptsidis, N.: New results on the linear complexity of binary sequences. IEEE Int. Symp. Inf. Theory, (2006) 2003–2007.
16. Limniotis, K., Kolokotronis, N., and Kalouptsidis, N.: Nonlinear complexity of binary sequences and connections with Lempel–Ziv compression. Sequences and Their Applications, Berlin, Germany: Springer-Verlag, **4086**, (2006) 168–179.
17. Limniotis, K., Kolokotronis, N., and Kalouptsidis, N.: On the linear complexity of sequences obtained by state-space generators. To be published in IEEE Trans. Inform. Theory, Apr. 2008.
18. MacWilliams, F. J., Sloane, N. J. A.: The Theory of Error Correcting Codes. Amsterdam, The Netherlands: North-Holland (1977).
19. Massey, J. L.: Shift register synthesis and BCH decoding. IEEE Trans. Inform. Theory **15** (1969) 122–127.
20. Massey, J. L., Serconek, S.: Linear complexity of periodic sequences: a general theory. in Proc. Advances in Cryptology - CRYPTO '96, Lecture Notes in Computer Science 358–371.
21. Matsui, M.: Linear cryptanalysis method for DES cipher. Advances in Cryptology - Eurocrypt '93 (Lecture Notes in Computer Science, Springer-Verlag) **765** (1993) 386–397.
22. Menezes, A. J., van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press (1996).
23. Niederreiter, H.: Some computable complexity measures for binary sequences. C. Ding, T. Helleseth, and H. Niederreiter, eds., in: Sequences and Their Applications, Discrete Mathematics and Theoretical Computer Science, Springer-Verlag (1999) 67–78.
24. Rizomiliotis, P., Kalouptsidis, N.: Results on the nonlinear span of binary sequences. IEEE Trans. Inform. Theory **51** (2005) 1555–1563.
25. Rizomiliotis, P., Kolokotronis, N., Kalouptsidis, N.: On the quadratic span of binary sequences. IEEE Trans. Inform. Theory **51** (2005) 1840–1848.
26. Rueppel, R. A.: Analysis and design of stream ciphers. Berlin, Germany: Springer-Verlang (1986).
27. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. IEEE Trans. Inform. Theory **24** (1978) 530–536.