Combining Certain Nonlinear Feedback Shift Registers

Berndt M. Gammel and Rainer Göttfert

Infineon Technologies AG
St. Martin-Str. 76
81541 Munich
Germany

berndt.gammel@infineon.com
rainer.goettfert@infineon.com

Abstract. Stream ciphers that deploy linear feedback shift registers (LFSRs) have been shown to be vulnerable under fast correlation attacks [20], [21], [14], algebraic attacks [7], [28], fast algebraic attacks [6], [1], and fault attacks [13]. We discuss certain nonlinear feedback shift registers (NLFSRs) recommended as substitutes for LFSRs in stream cipher systems.

1 Introduction

The period of the output sequence of an N-stage feedback shift register over the binary field $\mathbb{F}_2$ can be at most 2^N. N-stage shift register sequences whose periods attain the maximum value 2^N are called de Bruijn sequences. The number of translation distinct span-N de Bruijn sequences, and, consequently, the number of different N-stage feedback shift registers producing de Bruijn sequences is given by

$$Z(N) = 2^{2^{N-1}-N}, \tag{1}$$

as was shown by Flye Sainte-Marie [9] and de Bruijn [2], [3]. However, de Bruijn sequences have the drawback that the combination of several such sequences (combined by some Boolean combining function) does not lead to a sequence of larger period. This is quite obvious, considering that all sequences have periods that are powers of 2.

Therefore, the next best choice are N-stage shift register sequences of period $2^N - 1$ and $2^N - 2$. The fact that these sequences stem from an N-stage feedback shift register (have span N) and that their periods are $2^N - 1$ and $2^N - 2$, respectively, implies, in particular, an almost ideal k-tupel distribution over the full period for $1 \leq k \leq N$.

Definition 1. Let $N \geq 3$. An N-stage feedback shift register of type A is a nonlinear feedback shift register (NLFSR) of length N (i.e., comprising N delay elements) whose feedback function decomposes $\mathbb{F}_2^N$ into two cycles: one cycle contains all $2^N - 1$ nonzero vectors of $\mathbb{F}_2^N$ and the other cycle contains only the zero vector.

An N-stage feedback shift register of type B is a NLFSR whose feedback function is not affine and generates two cycles: the short cycle contains only the vector $(1, \ldots, 1) \in \mathbb{F}_2^N$, the long cycle contains all other vectors of $\mathbb{F}_2^N$.

An N-stage feedback shift register of type C is a NLFSR whose feedback function decomposes $\mathbb{F}_2^N$ into three cycles: one cycle contains the zero vector, one cycle the vector $(1, \ldots, 1) \in \mathbb{F}_2^N$, and the third cycle contains all the remaining vectors of $\mathbb{F}_2^N$.

Finally, an N-stage feedback shift register of type D is a NLFSR whose feedback function is not affine and generates two cycles: the short cycle contains the two vectors $(0, 1, 0, \ldots)$ and $(1, 0, 1, \ldots)$ of $\mathbb{F}_2^N$, the long cycle contains the remaining $2^N - 2$ vectors of $\mathbb{F}_2^N$.

By a *nontrivial* output sequence of a type A, B, C, or D shift register we mean any sequence produced by the shift register when any vector appearing in the corresponding long cycle is used to initialize the register. □

Table 1 contains some properties of N-stage NLFSRs of types A, B, C, and D. Notice that the NLFSRs of type C are the only ones that induce a cycle decomposition in $\mathbb{F}_2^N$ consisting of an odd number of cycles. It is known [11] that for such shift registers the corresponding feedback function always depends explicitly on all N variables. Thus, in order to minimize hardware costs, type-D shift registers are preferred over type-C registers.

In this paper we study the influence of a Boolean combining function on nontrivial output sequences of NLFSRs of type A, B, C, or D. Over the finite field $\mathbb{F}_2$ the effect of a Boolean combining function on individual sequences reduces to two simpler problems: termwise addition of sequences, and termwise multiplication of sequences. Since termwise addition is comparatively easy to analyze, we restrict ourselves to the study of termwise multiplication. It suffices to treat the product of two NLFSR sequences, as we can then proceed by induction to obtain results on the product of any finite number of sequences.

Any purely or ultimately periodic binary sequence σ possesses a unique minimal polynomial $m_\sigma \in \mathbb{F}_2[x]$. The minimal polynomial of σ contains a lot of information about σ:

NLFSRs of type	**A**	**B**	**C**	**D**
Length of shift register	N	N	N	N
Period of output sequence	$2^N - 1$	$2^N - 1$	$2^N - 2$	$2^N - 2$
Forbidden initializations	$(0, 0, \ldots, 0)$	$(1, 1, \ldots, 1)$	$(0, 0, \ldots, 0)$ $(1, 1, \ldots, 1)$	$(0, 1, 0 \ldots)$ $(1, 0, 1, \ldots)$
Linear complexity	$2^N - 2$	$2^N - 1$	$2^N - 2$	$2^N - 2$
Multiplicities of roots of minimal polynom	1	1	1 or 2	1 or 2
Degrees of irreducible factors of minimal polynomial divide	N	N	$N - 1$	$N - 1$
$x - 1$ divides the minimal polynomial	never	always	sometimes	sometimes
Feedback function contains constant term 1	no	yes	no	yes
Number of distinct cycles	2	2	3	2
Distribution of 0's and 1's in the full period	almost equidistributed	almost equidistributed	equidistributed	equidistributed
Sparse feedback functions exist	yes	yes	no	yes
Number of shift registers	$Z(N) - \frac{\varphi(2^N-1)}{N}$	$Z(N) - \frac{\varphi(2^N-1)}{N}$	$Z(N)$	less than $Z(N)$

Table 1. Properties of certain NLFSRs

1. The multiplicity of the element 0 as a root of m_σ equals the length of the preperiod of σ. In particular, σ is purely periodic if and only if $m_\sigma(0) \neq 0$.
2. The order of the polynomial m_σ coincides with the period of σ.
3. The polynomial m_σ is the characteristic polynomial of the shortest LFSR that can generate σ, so that the degree of m_σ is the linear complexity of σ, denoted by $L(\sigma)$.

We will derive formulæ for the minimal polynomial of the product of two NLFSR sequences produced by type A, B, C, or D shift registers. From the minimal polynomial one readily derives information on the period and linear complexity of the sequence. As all shift registers under consideration are nonsingular all sequences will be purely periodic.

In the theory of stream ciphers the determination of the linear complexity of the key stream is a fundamental problem. A good survey article on the theory of stream ciphers is Robshaw [23].

2 Results

This is a position paper. Proofs are omitted.

Proposition 1. *If σ is any nontrivial output sequence of an N-stage NLFSR of type A, then the minimal polynomial m_σ has the form*

$$m_\sigma = \prod_{i=1}^{r} f_i, \tag{2}$$

where all $f_i \in \mathbb{F}_2[x]$ are distinct irreducible polynomials with $\deg(f_i) \geq 2$ and $\deg(f_i)$ divides N.

Experimentally, we observed that the majority of type A shift register sequences σ have the minimal polynomial

$$m_\sigma(x) = x^{2^N-2} + \cdots + x^2 + x + 1 = \frac{x^{2^N} - x}{x(x-1)}.$$

NLFSRs of type A and B are closely related. The sequence $\tau = (t_n)_{n=0}^{\infty}$ is the output sequence of a type B shift register precisely if $\sigma = (s_n)_{n=0}^{\infty} = (t_n + 1)_{n=0}^{\infty}$ is the output sequence of a type A shift register. It follows that $m_\tau(x) = (x-1)m_\sigma(x)$, where m_σ has the form (2). The presence or absence of the factor $x - 1$ in the minimal polynomial has influence only on the behaviour of termwise addition but not on termwise multiplication.

Proposition 2. *Let σ be any nontrivial output sequence of an N-stage NLFSR of type C or D. Then the minimal polynomial of σ has the form*

$$m_\sigma = \prod_{i=1}^{r} f_i^{e_i}, \tag{3}$$

where $f_i \in \mathbb{F}_2[x]$ are distinct irreducible polynomials whose degrees divide $N-1$, $f_i(0) \neq 0$, and $1 \leq e_i \leq 2$ for $1 \leq i \leq r$. At least one exponent $e_i = 2$.

Theorem 1. *If $g(x) = x^N + c_{N-1}x^{N-1} + \cdots + c_1 x + c_0$ is a primitive polynomial over $\mathbb{F}_2$ of degree ≥ 3, then*

$$G(x_0, x_1, \ldots, x_{N-1}) = c_0 x_0 + c_1 x_1 \cdots + c_{N-1}x_{N-1} + x_1 x_2 \cdots x_{N-1}$$

is the feedback function of an N-stage NLFSR of type C. Any nontrivial initialization of the shift register results in an output sequence σ with the minimal polynomial

$$m_\sigma(x) = x^{2^N-2} - 1 = \prod f(x)^2, \tag{4}$$

where the product is extended over all irreducible polynomials $f \in \mathbb{F}_2[x]$ whose degrees divide $N-1$, with the exception of the polynomial $f(x) = x$.

We do not recommend the above theorem as a design rule for NLFSRs for cryptographic purposes. The output sequence behaves over a long part of the period like a LFSR sequence until, finally, the product term $x_1 \cdots x_{N-1}$ becomes effective. However, the theorem shows that for all $N \geq 3$, there is an N-stage NLFSR of type C whose nontrivial output sequences attain the maximum linear complexity value for such shift registers, namely $2^N - 2$.

The following facts are standard. See Selmer [27, Chap. 4]. Let $f, g \in \mathbb{F}_2[x]$ be nonconstant polynomials without multiple roots and not divisible by x. Then $f \vee g \in \mathbb{F}_2[x]$ is defined to be the polynomial whose roots are the distinct elements of the form $\alpha\beta$, where α is a root of f and β is a root of g. The polynomial $f \vee g$ is irreducible if and only if f and g are irreducible polynomials of relatively prime degrees. If $f \vee g$ is irreducible and $f(0)g(0) \neq 0$, then $\deg(f \vee g) = \deg(f)\deg(g)$. In this case, $f \vee g$ is the minimal polynomial of $\sigma\tau = (s_n t_n)_{n=0}^{\infty}$ whenever $\sigma = (s_n)_{n=0}^{\infty}$ is a periodic sequence with minimal polynomial f and $\tau = (t_n)_{n=0}^{\infty}$ is a periodic sequence with minimal polynomial g. Example: If $f(x) = x^2 + x + 1$ and $g(x) = x^3 + x + 1$, then $f(x) \vee g(x) = x^6 + x^4 + x^2 + x + 1$.

Theorem 2. *Let $\sigma = (s_n)_{n=0}^{\infty}$ and $\tau = (t_n)_{n=0}^{\infty}$ be nontrivial output sequences of an M-stage and N-stage NLFSR, respectively, of type A or B. Assume that the lengths of the shift registers are relatively prime, i.e.,* $\gcd(M, N) = 1$. *If $m_\sigma = \prod_{i=1}^{r} f_i$ and $m_\tau = \prod_{j=1}^{s} g_j$, then the product sequence $\sigma\tau = (s_n t_n)_{n=0}^{\infty}$ has the minimal polynomial*

$$m_{\sigma\tau} = \prod_{\substack{1 \le i \le r \\ 1 \le j \le s}} (f_i \vee g_j). \tag{5}$$

In fact, (5) *is the canonical factorization of $m_{\sigma\tau}$.*

Comparing the degrees of both sides in (5), using $\deg(f_i \vee g_j) = \deg(f_i)\deg(g_j)$, yields

$$L(\sigma\tau) = L(\sigma)L(\tau). \tag{6}$$

This result on the linear complexity of the product sequence carries over to any finite number of NLFSR sequences of type A or B, provided that the lengths of the corresponding shift registers are pairwise relatively prime. Equation (6) can also be derived from [10], where it is shown that for purely periodic sequences $\sigma_1, \ldots, \sigma_k$ with pairwise relatively prime periods the linear complexity of the product sequence $\omega = \sigma_1 \cdots \sigma_k$ equals the product of the linear complexities of the individual sequences. As a matter of fact, two positive integers a and b are relatively prime if and only if $2^a - 1$ and $2^b - 1$ are. Thus, our assumption that the lengths of the NLFSRs be pairwise relatively prime implies that the periods are pairwise relatively prime as well, so that the required premise in [10] is true.

The situation is different if we work in the general finite field $\mathbb{F}_q$. The counterparts of the N-stage NLFSRs of type A considered here are now N-stage NLFSRs over $\mathbb{F}_q$ whose nontrivial output sequences have period $q^N - 1$. To obtain reasonable lower bounds on the linear complexity of the product of such NLFSR sequences in $\mathbb{F}_q$, we need again the requirement that the lengths of the shift registers be relatively prime. However, for any prime power q with $q \neq 2$, the numbers $q^a - 1$ and $q^b - 1$ are never relatively prime, no matter which positive integers a and b are used.

The next theorem is concerned with the multiplication of one NLFSR sequence of type A or B with another one of type C or D.

Theorem 3. *Let σ be a nontrivial output sequence of an M-stage NLFSR of type A or B, and let τ be a nontrivial output sequence of an N-stage NLFSR of type C or D. Assume that* $\gcd(M, N-1) = 1$. *If $m_\sigma = \prod_{i=1}^{r} f_i$*

and $m_\tau = \prod_{j=1}^{s} g_j^{e_j}$ *are the canonical factorizations of the minimal polynomials of* σ *and* τ*, respectively, then the canonical factorization of the minimal polynomial of the product sequence* $\sigma\tau$ *is given by*

$$m_{\sigma\tau} = \prod_{\substack{1\le i\le r\\ 1\le j\le s}} (f_i \vee g_j)^{e_j}. \tag{7}$$

While the analysis of the combinational behaviour of NLFSRs of type A is surprisingly easy, NLFSRs of type C and D are more challenging. The periods of the sequences are no longer relatively prime. According to Proposition 2, the minimal polynoms have multiple roots, so that the results of Herlestam [12] do not apply. Furthermore, there will be coincidences of root products of the same multiplicity, so that the well established root-counting method (see [15], [22]) fails to lead to lower bounds on the linear complexity of the product sequence in this case. Nonetheless, using the apparatus of generating functions lower bounds can be derived.

We recall some well known facts from the theory of linear recurring sequences. See Lidl and Niederreiter [18, Chap. 8].

Let f be a polynomial over $\mathbb{F}_2$ with $\deg(f) = d \ge 1$. We define $M(f)$ to be the set of all periodic binary sequences with minimal polynomial f. By $S(f)$ we denote the set of all binary sequences with characteristic polynomial f. Under termwise operations, $S(f)$ is a vector space over $\mathbb{F}_2$ of dimension d. Clearly, $M(f) \subseteq S(f)$. If f is irreducible, then $S(f) = M(f) \cup \{\mathbf{0}\}$, where $\mathbf{0} = (0, 0, \dots)$ is the zero sequence.

There are two fundamental linear operators on $\mathbb{F}_2^\infty$. The shift operator T, defined by $T\sigma = (s_{n+1})_{n=0}^\infty$, and the decimation operator D, defined by $D\sigma = (s_{2n})_{n=0}^\infty$, for all binary sequences $\sigma = (s_n)_{n=0}^\infty$. If f is any nonzero polynomial over $\mathbb{F}_2$, then $f(T)$ is again a linear operator on $\mathbb{F}_2^\infty$, and $S(f) = \{\sigma \in \mathbb{F}_2^\infty : f(T)\sigma = \mathbf{0}\}$. It is well known and easy to show that $S(f)$ is closed under the actions of T and D, that is, $T\sigma \in S(f)$ and $D\sigma \in S(f)$ whenever $\sigma \in S(f)$.

If σ is any periodic binary sequence, then we denote the minimal polynomial of the decimated sequence $D\sigma$ by $m_{D\sigma}$. The minimal polynomial of the shifted sequence $T\sigma$ is denoted by $m_{T\sigma}$.

If f and g are binary polynomials with $f = g^2$, then we also write $g = \sqrt{f}$. Consider the canonical factorization of the binary polynomial

$$h = h_1^2 \cdots h_c^2 h_{c+1} \cdots h_{c+m},$$

where the h_i, $1 \le i \le c+m$, are distinct irreducible polynomials over $\mathbb{F}_2$. The *squarefree part* of h is denoted by $\langle h \rangle$, defined by $\langle h \rangle = h_1 \cdots h_{c+m}$,

and can be computed by $\langle h \rangle = h/\sqrt{\gcd(h, h')}$, where h' is the first derivative of h. The *quadratic part* of the polynomial h is $\gcd(h, h') = h_1^2 \cdots h_c^2$.

We now deal with the case where both sequences σ and τ are NLFSR sequences of type C or D. Let M and N be the lengths of the shift registers producing σ and τ. According to Proposition 2, the minimal polynomial of σ has the form

$$m_\sigma = f_1^2 \cdots f_a^2 f_{a+1} \cdots f_{a+k}, \tag{8}$$

where the f_i are distinct irreducible binary polynomials whose degrees divide $M-1$ and with $f_i(x) \neq x$ for $1 \leq i \leq a+k$, $a \geq 1$, and $k \geq 0$. Likewise,

$$m_\tau = g_1^2 \cdots g_b^2 g_{b+1} \cdots g_{b+l}, \tag{9}$$

where all $g_j \in \mathbb{F}_2[x]$ are distinct, irreducible, not equal to x, and with $\deg(g_j)$ dividing $N-1$, $b \geq 1$, and $l \geq 0$.

Theorem 4. *Let σ and τ be nontrivial output sequences of an M-stage and N-stage NLFSR, respectively, of type C or D. Assume that* $\gcd(M-1, N-1) = 1$. *Let the minimal polynomial of σ and τ be given by* (8) *and* (9). *Then the minimal polynomial of the product sequence $\sigma\tau$ is*

$$m_{\sigma\tau} = \frac{\displaystyle\prod_{i=1}^{a+k} \prod_{j=1}^{b+l} (f_i \vee g_j)^2}{\displaystyle\prod_{i=a+1}^{a+k} \prod_{j=b+1}^{b+l} (f_i \vee g_j) \prod_{(i,j) \in I_0 \times J_0} (f_i \vee g_j)^2 \prod_{(i,j) \in I_1 \times J_1} (f_i \vee g_j)^2}.$$

The index sets I_0 and I_1 appearing in the formula are disjoint subsets of $\{1, \ldots, a\}$, whereas J_0 and J_1 are disjoint subsets of $\{1, \ldots, b\}$. The index sets, respectively the last two products in the denominator, can be computed from the sequences σ and τ as follows: First, determine the sequences $D\sigma$, $D\tau$, $DT\sigma$, $DT\tau$, and their minimal polynomials. Next, compute the binary polynomials

$$S_0 = \frac{\langle m_\sigma \rangle}{m_{D\sigma}}, \quad S_1 = \frac{\langle m_\sigma \rangle}{m_{DT\sigma}}, \quad T_0 = \frac{\langle m_\tau \rangle}{m_{D\tau}}, \quad T_1 = \frac{\langle m_\tau \rangle}{m_{DT\tau}}. \tag{10}$$

Finally, compute the canonical factorizations of the polynomials in $\mathbb{F}_2[x]$:

$$S_0 = \prod_{i \in I_0} f_i, \quad S_1 = \prod_{i \in I_1} f_i, \quad T_0 = \prod_{j \in J_0} g_j, \quad T_1 = \prod_{j \in J_1} g_j. \tag{11}$$

This defines the index sets I_0, I_1, J_0, and J_1. From here, the last two products in the denominator of the big fraction can be derived.

By considering the degree of $m_{\sigma\tau}$, we obtain a formula for the linear complexity of the product sequence $\sigma\tau$.

Corollary 1. *Under the provisions of Theorem 4 we have*

$$\begin{aligned} L(\sigma\tau) =& L(\sigma)L(\tau) - 2\deg\left(\sqrt{\gcd(m_\sigma, m'_\sigma)}\right)\deg\left(\sqrt{\gcd(m_\tau, m'_\tau)}\right) \\ & - 2\deg(S_0)\deg(T_0) - 2\deg(S_1)\deg(T_1), \end{aligned}$$

where S_0, T_0, S_1, and T_1 are the polynomials in (10).

Notice that for the calculation of the linear complexity of $\sigma\tau$, we do not need to know the canonical factorizations of m_σ or m_τ. The required minimal polynomials m_σ, m_τ, $m_{D\sigma}$, etc., can be assessed by the Berlekamp-Massey algorithm. However, given the fact that the best known upper bounds for the linear complexities of the sequences are the periods of the sequences, the formula of Laksov [16, Lemma 3] might be more favourable.

It is interesting to note that the costs for retrieving a NLFSR of the types considered here by a random computer search are about the same as the costs for computing the minimal polynomial associated with the shift register. Both tasks are feasible on a standard workstation for shift register lengths up to $N = 30$. (By the minimal polynomial associated with the shift register we mean, of course, the uniquely determined minimal polynomial of any nontrivial output sequence of the shift register.)

Our proof of Theorem 4 proceeds through a series of lemmas. We state just one of the lemmas, the most crucial one. In the lemma, (n) denotes the binary sequence $(n)_{n=0}^{\infty} = (0, 1, 0, 1, \dots)$. Likewise, $(n+1) = (1, 0, 1, 0, \dots)$.

Lemma 1. *Let f and g be irreducible polynomials over $\mathbb{F}_2$ of relatively prime degrees none of them equal to x. For any $\sigma_1 \in S(f)$, $\sigma_2 \in M(f)$, $\tau_1 \in S(g)$, and $\tau_2 \in M(g)$, the sequence*

$$\sigma_1\tau_1 + (n)\sigma_2\tau_1 + (n+1)\sigma_1\tau_2 \tag{12}$$

is either the zero sequence or has minimal polynomial $(f \vee g)^2$. The sequence is the zero sequence if and only if ($\sigma_1 = \mathbf{0}$ and $\tau_1 = \mathbf{0}$) or ($\sigma_1 = \sigma_2$ and $\tau_1 = \tau_2$).

Definition 2. Let σ be a nontrivial output sequence of an N-stage NLFSR of type C or D. Let $m_\sigma = f_1^2 \cdots f_a^2 f_{a+1} \cdots f_{a+k}$. Then σ has a unique representation of the form

$$\sigma = \sum_{i=1}^{a} [\sigma_i^{(0)} + (n)\sigma_i^{(1)}] + \sum_{i=a+1}^{a+k} \sigma_i \tag{13}$$

with $\sigma_i^{(0)} \in S(f_i)$, $\sigma_i^{(1)} \in M(f_i)$ for $1 \le i \le a$, and $\sigma_i \in M(f_i)$ for $a+1 \le i \le a+k$. We call the underlying NLFSR *friendly* if for all $i = 1, \ldots, a$ we have $\sigma_i^{(0)} \neq \mathbf{0}$ and $\sigma_i^{(0)} \neq \sigma_i^{(1)}$. We call the NLFSR *almost friendly* if the requirement is fulfilled for all but one index i. □

Recall Theorem 4 for a motivation of the above definition. The two index sets I_0 and I_1 in Theorem 4 are empty if and only if the underlying NLFSR that produces the sequence σ is friendly.

The property of being friendly or almost friendly is independent of the particular (nontrivial) initialization of the type C or D shift register. The idea behind the definition of *almost friendly* is the following: the minimal polynomial m_σ of a nontrivial output sequence of a type C or D shift register will often contain the factor $(x-1)^2$. As soon as this happens the NLFSR can not be friendly anymore, as then, necessarily, in (13) the sequence (n) or $(n+1)$ will occur in the first sum. That means that there is one index i between 1 and a for which, either $\sigma_i^{(0)}$ is the zero sequence, or the sequences $\sigma_i^{(0)}$ and $\sigma_i^{(1)}$ are identical with the sequence all of whose terms are 1.

In a heuristic approach, we treat the sequences $\sigma_i^{(0)}$ in (13) as if they were random objects. At this we can show that the probability for the underlying NLFSR to be friendly or almost friendly converges to 1 as N goes to infinity. For instance, for $N = 28$, we obtain the probability 0.93.

Theorem 5. *Under the provisions of Theorem 4 and the additional assumption that at least one of the underlying NLFSRs is friendly, we have*

$$m_{\sigma\tau} = \prod_{\substack{1 \le i \le a \\ 1 \le j \le b+l}} (f_i \vee g_j)^2 \prod_{\substack{a+1 \le i \le a+k \\ 1 \le j \le b}} (f_i \vee g_j)^2 \prod_{\substack{a+1 \le i \le a+k \\ b+1 \le j \le b+l}} (f_i \vee g_j),$$

and, consequently,

$$L(\sigma\tau) = L(\sigma)L(\tau) - 2\deg\left(\sqrt{\gcd(m_\sigma, m'_\sigma)}\right)\deg\left(\sqrt{\gcd(m_\tau, m'_\tau)}\right).$$

If $k = l = 0$, or, equivalently, if all factors in (8) *and* (9) *have multiplicity* 2*, then*

$$L(\sigma\tau) = \frac{1}{2}L(\sigma)L(\tau).$$

The next corollary is of practical importance.

Corollary 2. *Let σ be the output sequence of an almost friendly M-stage NLFSR of type C or D. Let τ be the output sequence of an almost friendly*

N-stage NLFSR of the same type. If $L(\sigma) = 2^M - 2$, $L(\tau) = 2^N - 2$, *and* $\gcd(M-1, N-1) = 1$, *then*

$$\frac{1}{2}\left(2^M - 2\right)\left(2^N - 2\right) - 2 \le L(\sigma\tau) \le \frac{1}{2}\left(2^M - 2\right)\left(2^N - 2\right).$$

Our experimental investigations suggest that the specified conditions in Corollary 2 are reasonable assumptions, fulfilled by the majority of type-C and D shift registers. Here is an algorithm to find out whether or not a NLFSR of type C or D is friendly.

Algorithm.
Given: a NLFSR of type C or D.

1. Produce a nontrivial output sequence $\sigma = (s_n)_{n=0}^{\infty}$ of the shift register.
2. Compute the minimal polynomial m_σ.
3. Compute the squarefree part $\langle m_\sigma \rangle$ of m_σ.
4. Apply the decimation operator D to σ: $D\sigma = (s_{2n})_{n=0}^{\infty}$.
5. Compute the linear complexity $L(D\sigma)$ of the decimated sequence $D\sigma$.
6. If $L(D\sigma) = \deg(\langle m_\sigma \rangle)$, then all $\sigma_i^{(0)} \neq \mathbf{0}$ in (13).
7. Apply the shift operator T to σ: $T\sigma = (s_{n+1})_{n=0}^{\infty}$.
8. Apply the decimation operator to $T\sigma$ to produce $DT\sigma$.
9. Compute the linear complexity $L(DT\sigma)$ of the sequence $DT\sigma$.
10. If $L(DT\sigma) = \deg(\langle m_\sigma \rangle)$, then $\sigma_i^{(0)} \neq \sigma_i^{(1)}$ for all $1 \le i \le a$ in (13).

Answer: If the premises in (6) and (10) are true then the given NLFSR is friendly.

The formulæ for the minimal polynomial of the product of two NLFSR sequences presented above generalize to formulæ for the product of any finite number of sequences under similar restrictions regarding the lengths of the shift registers. Results on the termwise addition of sequences are easily derived. Combining the results on addition and multiplication, we obtain formulæ for the minimal polynomial of sequences of the form $\omega = F(\sigma_1, \ldots, \sigma_k)$, where F is any boolean combining function defined on $\mathbb{F}_2^k$. If none of the individual sequences has a minimal polynomial containing the factor $x - 1$, then the minimal polynomial of ω is uniquely determined. Otherwise, it may range over some relatively small range. By considering the degrees of the minimal polynomials, we obtain the following two corollaries.

Corollary 3. *Let* $F : \mathbb{F}_2^k \to \mathbb{F}_2$ *be an arbitrary boolean function given in its algebraic normal form* $F(x_1, \ldots, x_k)$, *where* $k \ge 1$. *For each* $i =$

$1, \ldots, k$, *let* σ_i *be the nontrivial output sequence of an* N_i*-stage NLFSR of type A. If* $\gcd(N_i, N_j) = 1$ *for* $1 \leq i \neq j \leq k$*, then the linear complexity of the sequence* $\omega = F(\sigma_1, \ldots, \sigma_k)$ *is*

$$L(\omega) = F(L(\sigma_1), \ldots, L(\sigma_k)) \tag{14}$$

with the understanding that in (14)*, the polynomial* F *is evaluated over the integers.*

A formula of the type in (14) was proposed by Rueppel and Staffelbach [26, Theorem 3] in the context of LFSR sequences of maximum period length. They showed that the claim already holds if the lengths N_i of the LFSRs are distinct, they do not necessarily have to be pairwise relatively prime like in the nonlinear case.

An alternative proof of Corollary 3 can be derived from Golić [10, Theorem 5] using the fact that for positive integers a and b the statements $\gcd(a, b) = 1$ and $\gcd(2^a - 1, 2^b - 1) = 1$ are equivalent.

The first paper that presented a compact formula for the linear complexity of combined linear feedback shift register sequences is Brynielsson [4].

Corollary 4. *Let* $F : \mathbb{F}_2^k \to \mathbb{F}_2$ *with* $k \geq 1$ *be an arbitrary boolean function and* $F(x_1, \ldots, x_k)$ *its algebraic normal form. For each* $i = 1, \ldots, k$*, let* σ_i *be the nontrivial output sequence of some friendly* N_i*-stage NLFSR of type C or D. Assume that* $\gcd(N_i - 1, N_j - 1) = 1$ *for* $1 \leq i \neq j \leq k$*. If for each* $i = 1, \ldots, k$ *the minimal polynomial of* σ_i *contains only multiple roots, then the linear complexity of the sequence* $\omega = F(\sigma_1, \ldots, \sigma_k)$ *is*

$$L(\omega) = 2F\left(\frac{L(\sigma_1)}{2}, \ldots, \frac{L(\sigma_k)}{2}\right), \tag{15}$$

where in (15)*, the polynomial* F *is evaluated over the integers.*

A crucial question for the practical relevance of the results presented so far is, of course, the actual availability of the discussed NLFSRs. The situation is not as convenient as in the linear case, where one only has to grab some primitive polynomial from $\mathbb{F}_2[x]$ in order to get a reasonable LFSR. Instead, the NLFSRs discussed here must be found by computer search. Comparing the number $Z(N)$ in equation (1) with the total number of nonsingular binary N-stage feedback shift registers given by 2^E with $E = 2^{N-1}$, one sees that the odds to find "a good" N-stage NLFSR by a blind random search are roughly $1 : 2^N$.

Only a modest amount of theory is available, at least in the open literature, that can be used to speed up the search process or narrow down the search domain to some extent. For instance, one can apply results concerning the binary complementation and cycle reversal of NLFSR sequences as described in Walker [29] and similar things.

Other methods than pure computer search are of high interest. Abraham Lempel derived in [17] a recursive formula for the feedback function of a shift register producing de Bruijn sequences. From there, it does not take a genius to arrive at the following result.

Theorem 6. *Let $F(x_0, x_1, \ldots, x_{N-1})$ be the feedback function of some N-stage NLFSR of type A. Then for any $j \in \{1, 2, \ldots, N\}$*

$$G_j(x_0, x_1, \ldots, x_N) = x_N + F(x_0 + x_1, \ldots, x_{N-1} + x_N) + \prod_{\substack{i=1 \\ i \neq j}}^{N} x_i$$

is the feedback function of an $(N+1)$-stage NLFSR of type A.

A few remarks on the linear complexity of the output sequences of the NLFRSs considered in this paper are due. For de Bruijn sequences produced by an N-stage NLFSR a lower bound for the linear complexity is known to be $2^{N-1} + N$. See Chan, Games, and Key [5]. It seems that no nontrivial lower bounds on the linear complexities of the NLFSR sequences considered here are known. We have carried out extensive computer calculations to satisfy our curiosity. It turned out that the maximum possible value for the linear complexity for each shift register type, as displayed in Table 1, is also the typical value for the linear complexity.

In analogy to the quoted lower bound for the linear complexity of de Bruijn sequences one is inclined to anticipate a similar lower bound for the linear complexity of the NLFSR sequences treated in this paper. However, this is misleading. We examined more than 100 million NLFSRs of type A of lengths varying between $N = 4$ and $N = 30$, and among them, we did encounter 28 NLFSRs whose nontrivial output sequences had a linear complexity smaller than half the period length.

It was conjectered by Rueppel [24], [25], and confirmed by Dai and Yang [8], and by Meidl and Niederreiter [19] that the linear complexity of a periodic random sequence (a randomly generated finite bit string that is then repeated ad infinitum) is close to the period length. Thus our experiments indicate that the nontrivial output sequences of the NLFSRs

considered in this paper typically share their linear complexity behaviour with true periodic random sequences.

References

1. F. Armknecht: Improving fast algebraic attacks, FSE 2004, Delhi, India, 2004.
2. N. G. de Bruijn: A combinatorial problem, *Indag. Math.* **8**, 461–467 (1946).
3. N. G. de Bruijn: Acknowledgement of priority to C. Flye Sainte-Marie on the counting of 2^n zeros and ones that show each n-letter word exactly once, Technical Report TH 75-WSK-06, Technische Hogeschool Eindhoven, 1975.
4. L. Brynielsson: On the linear complexity of combined shift register sequences, *Advances in Cryptology – EUROCRYPT '85* (F. Pichler, ed.), Lecture Notes in Computer Science, vol. 219, pp. 156–160, Springer-Verlag, 1985.
5. A. H. Chan, R. A. Games, and E. L. Key: On the complexities of de Bruijn sequences, *J. Combin. Theory Ser. A* **33**, 233–246 (1982).
6. N. Courtois: Fast algebraic attacks on stream ciphers with linear feedback, *Advances in Cryptology – CRYPT0 2003* (D. Boneh, ed.), Lecture Notes in Computer Science, vol. 2729, pp. 176–194, Springer-Verlag, 2003.
7. N. Courtois and W. Meier: Algebraic attacks on stream ciphers with linear feedback, *Advances in Cryptology – EUROCRYPT 2003* (E. Biham, ed.), Lecture Notes in Computer Science, vol. 2656, pp. 345–359, Springer-Verlag, 2003.
8. Z.-D. Dai and J.-H. Yang: Linear complexity of periodically repeated random sequences, *Advances in Cryptology – EUROCRYPT '91* (D. W. Davies, ed.), Lecture Notes in Computer Science, vol. 547, pp. 168–175, Springer-Verlag, 1991.
9. C. Flye Sainte-Marie: *L' Interméd. Math.* **1**, 107–110 (1894).
10. J. Dj. Golić: On the linear complexity of functions of periodic GF(q) sequences, *IEEE Trans. Inform. Theory* **35**, 69–75 (1989).
11. S. W. Golomb: *Shift Register Sequences*, Aegean Park Press, Laguna Hills, Cal., 1982.
12. T. Herlestam: On functions of linear shift register sequences, *Advances in Cryptology – EUROCRYPT '85* (F. Pichler, ed.), Lecture Notes in Computer Science, vol. 219, pp. 119–129, Springer-Verlag, 1986.
13. J. H. Hoch and A. Shamir: Fault analysis of stream ciphers, *CHES 2004* (M. Joye and J.-J. Quisquater, eds.) Lecture Notes in Computer Science, vol. 3156, pp. 240–253, Springer-Verlag, 2004.
14. T. Johansson and F. Jönsson: Fast correlation attacks based on turbo code techniques, *Advances in Cryptology – CRYPTO '99* (M. Wiener, ed.), Lecture Notes in Computer Science, vol. 1666, pp. 181–197, Springer-Verlag, 1999.
15. E. Key: An analysis of the structure and complexity of nonlinear binary sequence generators, *IEEE Trans. Inform. Theory* **IT-22**, 732–736 (1976).
16. D. Laksov: Linear recurring sequences over finite fields, *Math. Scand.* **16**, 181-196 (1965).
17. A. Lempel: On a homomorphism of the de Bruijn graph and its applications to the design of feedback shift registers, *IEEE Trans. Computers* **C-19**, 1204–1209 (1970).
18. R. Lidl and H. Niederreiter: *Finite Fields*, Encyclopedia of Mathematics and Its Applications, vol. 20, Addison-Wesley, Reading, Mass., 1983. (Now Cambridge Univ. Press.)

19. W. Meidl and H. Niederreiter: On the expected value of the linear complexity and the k-error linear complexity of periodic sequences, *IEEE Trans. Inform. Theory* **48**, 2817–2825 (2002).
20. W. Meier and O. Staffelbach: Fast correlation attacks on stream ciphers, *Advances in Cryptology – EUROCRYPT '88* (C. Günther, ed.), Lecture Notes in Computer Science, vol. 330, pp. 301–314, Springer-Verlag, 1988.
21. W. Meier and O. Staffelbach: Fast correlation attacks on certain stream ciphers, *J. Cryptology* **1**, 159–176 (1989).
22. K. G. Paterson: Root counting, the DFT and the linear complexity of nonlinear filtering, *Designs, Codes and Cryptography* **14**, 247–259 (1998).
23. M. J. B. Robshaw: Stream Ciphers, RSA Laboratories Technical Report TR-701, RSA Labs., Redwood City, CA, 1995.
24. R. A. Rueppel: *New Approaches to Stream Ciphers*, Ph.D. Thesis, Swiss Federal Institute of Technology, Zürich, 1984.
25. R. A. Rueppel: Linear complexity and random sequences, *Advances in Cryptology – EUROCRYPT '85* (F. Pichler, ed.), Lecture Notes in Computer Science, vol. 219, pp. 167–188, Springer-Verlag, 1985.
26. R. A. Rueppel and O. J. Staffelbach: Products of linear recurring sequences with maximum complexity, *IEEE Trans. Inform. Theory* **IT-33**, 124–131 (1987).
27. E. S. Selmer: *Linear Recurrence Relations over Finite Fields*, Department of Mathematics, Univ. of Bergen, 1966.
28. A. Shamir, J. Patarin, N. Courtois, and A. Klimov: Efficient algorithms for solving overdefined systems of multivariate polynomial equations, *Advances in Cryptology – EUROCRYPT 2000* (B. Preneel, ed.), Lecture Notes in Computer Science, vol. 1807, pp. 392–407, Springer-Verlag, 2000.
29. E. Walker: Non-linear recursive sequences, *Can. J. Math.* **11**, 370-378 (1959).