Powers of Subfield Polynomials and Algebraic Attacks on Word-Based Stream Ciphers

Sondre Rønjom

Nasjonal sikkerhetsmyndighet
Oslo, Norway
sondrer@gmail.com

Abstract. In this paper we investigate univariate algebraic attacks on filter generators over extension fields $\mathbb{F}_q = \mathbb{F}_{2^n}$ with focus on the Welch-Gong (WG) family of stream ciphers. Our main contribution is to break WG-5, WG-7, WG-8 and WG-16 by combining results on the so-called spectral immunity (minimum distance of certain cyclic codes) with properties of the WG type stream cipher construction. The spectral immunity is the univariate analog of algebraic immunity and instead of measuring degree of multiples of a multivariate polynomial, it measures the minimum number of nonzero coefficients of a multiple of a univariate polynomial. Based on the structure of the general WG-construction, we deduce better bounds for the spectral immunity and the univariate analog of algebraic attacks.

1 Introduction

There exist at least five published variants of the WG construction; WG-5 [15], WG-7 [16], WG-8 [17], WG-16 [18] and WG-29 [19]. In this section we present basic results that make up the machinery of these constructions, needed for our cryptanalysis of WG-ciphers in Section 3.

1.1 M-sequences, Unitary Sequences and Linear Complexity

For a much better introduction to the relationship between finite fields and sequences, the reader is referred to [1] and [2]. Let $\mathbb{F}_{q^n}$ denote a n-th order extension of the binary field $\mathbb{F}_q$ of $q = 2^k$ elements, defined by a polynomial $m(x)$ over $\mathbb{F}_q$. In order to simplify the presentation we assume that $m(x)$ is a primitive polynomial. The polynomial $m(x)$ defines a linear feedback shift register (LFSR) over $\mathbb{F}_q$ of length n that generates a maximal sequence (or *m-sequence*) of period $q^n - 1$; if initzialised in a non-zero state the LFSR spans the coefficient vectors of exactly the elements of the multiplicative group $\mathbb{F}_{q^n}^*$. Moreover, if the LFSR is initizialised in a state $S_0 = (s_0, s_1, \ldots, s_{n-1}) \in \mathbb{F}_q^n$, the LFSR generates a sequence obeying the recurrence relation

$$s_{t+n} = s_t c_0 + s_{t+1} c_1 + \ldots + s_{t+n-1} c_{n-1}. \quad (1)$$

defined by the coefficients of $m(x)$. The minimal polynomial of a periodic sequence s over $\mathbb{F}_{q^n}$ is the polynomial of least degree generating that sequence. The degree of this polynomial is what is called the *linear complexity* of the sequence. Let $\alpha \in \mathbb{F}_{q^n}$ be, for sake of simplicity, a primitive element. For $\beta \in \mathbb{F}_{q^n}^*$ and $d \in \{0, 1, \ldots, q^n - 1\}$ we call the sequence

$$b_{d,t} = \beta(\alpha^t)^d, t = 0, 1, 2, \ldots$$

over $\mathbb{F}_{q^n}$ a *unitary sequence.* Unitary sequences $b_{d,t}$ are the simplest forms of nonzero sequences in the sense that their linear complexity is 1, since their minimal polynomials are the linear polynomials $x + \alpha^d$. It is well-known that the minimal polynomial of the sum of two sequences a_t and b_t is equal to the least common multiple of their individual minimal polynomials. Thus the sum of two unitary sequences $a_t = \beta_1(x\alpha^t)^{d_1}$ and $b_t = \beta_2(x\alpha^t)^{d_2}$ has simply minimal polynomial $m(x) = (x + \alpha^{d_1})(x + \alpha^{d_2})$. In general, if I is a random distinct subset of $\{0, 1, 2, \ldots, q^n - 1\}$ and c_i are random nonzero constants of $\mathbb{F}_{q^n}$, the polynomial

$$P(x) = \sum_{i \in I} c_i x^i$$

defines a sum of unitary sequences $z_t = \sum_{i \in I} c_i(x\alpha^t)^i$ with minimal polynomial $m(x) = \prod_{i \in I}(x + \alpha^i)$ and linear complexity $|I|$.

2 Filter Generators and Algebraic Attacks over $\mathbb{F}_{2^n}$

Filter generators have been well-studied in literature and consists usually of a binary m-sequence generating LFSR of length n, a Boolean function f in k variables and a subset of tapping positions $I \subset \{i_1, i_2, \ldots, i_k\} \subset \{0, 1, 2, \ldots, n\}$. In this section we quickly recapture the current state of algebraic attacks on such constructions, but in terms of univariate polynomial equations. In the rest of the paper, all operations on polynomials over $\mathbb{F}_{q^n}$ are modulo $x^{q^n} + x$. Let $L(x) = \sum_{i=0}^{n-1} x^{2^i}$ denote the trace from $\mathbb{F}_q = \mathbb{F}_{2^n}$ to $\mathbb{F}_2$ and $\alpha \in \mathbb{F}_{2^n}$ a root of the LFSR feedback polynomial. Since the shift-register obeys a linear recursion, each bit s_{t+i} of the state S_t at time t can be described linearly by $L_{t+i}(x) = L(x\alpha^{t+i})$ where x is the initial state. If the state of the LFSR at time t is $S_t = (s_t, s_{t+1}, \ldots, s_{t+n-1})$, a binary keystream sequence can be generated by

$$\begin{aligned} z_t =& f(s_{t+i_1}, s_{t+i_2}, \ldots, s_{t+i_k}) \\ =& f(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x)) \end{aligned}$$

The bits of the sequence z_t are successively exored with the plaintext bits to form a ciphertext sequence. The choice of LFSR, tapping positions and Boolean function all have various effects on the cryptographic quality of the resulting keystream z_t.

2.1 Algebraic Attacks

In algebraic cryptanalysis (see for instance [6] and [5]) of a filter generator the adversary tries to solve an associated equation system relating unknown state-variables with keystream values. Then if the adversary has observed a sequence of keystream bits beginning at time t, $(z_t, z_{t+1}, \ldots, z_{t+m})$, she can set up a system of equations of the form

$$\begin{aligned} z_t =& f(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x)) \\ =& F_t(x), t = 0, 1, 2, \ldots. \end{aligned}$$

The Boolean function f contains monomials in n variables of degree up to $d = \deg(f)$, thus the univariate polynomial $F_t(x)$ can have at most $D = \sum_{i=0}^{d} \binom{n}{i}$ nonzero coefficients (exactly those x^i where $wt(i) \leq d$). This means that if the adversary observes D keystream bits, she can set up a system of at most D equations in D unknowns over $\mathbb{F}_{2^n}$ and solve using linear algebra. In multivariate cryptanalysis the complexity is given by $O(D^{log_2(7)})$. Notice that $F_{t+i}(x) = F_t(x\alpha^i)$ and that the coefficients of $F_t(x) = \sum_{wt(i)\leq d} c_i x^i \alpha^{ti}$ span cyclic vectors of the form

$$v_t = (\alpha^{ti_1}, \alpha^{ti_2}, \ldots, \alpha^{ti_D}).$$

If we let D such vectors for $t = 0, 1, 2, \ldots, D - 1$ span a $D \times D$ matrix M, the resulting matrix is a Vandermonde matrix and can be manipulated more efficiently than generic matrices (the inverse can be computed in $O(Dlog(D)^2)$). Moreover, if $X = (c_{i_1}x^{i_1}, c_{i_2}x^{i_2}, \ldots, c_{i_D}x^{i_D})$ then $M \cdot X = (z_0, z_1, \ldots, z_{D-1})$ and

$$M^{-1}(z_0, z_1, \ldots, z_{D-1}) = (x_{i_0}, x_{i_1}, \ldots, x_{i_D})$$

If we compute X from z, we can easily recover the initial state x from one of the equations $c_{i_j}x^{i_j} = x_{i_j}$. In practice one can pre-compute one of the columns of the inverse to recover x from a pre-chosen value x_{i_j}. This is essentially the improved algebraic attack presented in [14].

2.2 Algebraic Attacks and Low-Degree Polynomials

An often more keystream efficient method is to make use of low-degree multivariate multiples of f and $f + 1$. Moreover, if there exist a multivariate Boolean polynomial g in the ideal spanned by f over $\mathbb{F}_2$ of lower degree e, the adversary can use the relation

$$g(S_t)(z_t + f(S_t)) = 0$$

which yields a new valid equation each time $z_t = 0$ since the zeros of f is a subset of any multiple g. If we let $G_t(x) = g(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x))$, we can construct a system of equations

$$G_{t_i}(x) = 0$$

for all t_i when $z_{t_i} = 0$. Let $T = \{t_1, t_2, \ldots, t_E\}$ where $E = \sum_{i=0}^{e} \binom{n}{i}$. The equations involve at most E nonzero coefficients so we can set up a $E \times E$ matrix M spanned by coefficient vectors $v_{t_j} = (\alpha^{t_j i_1}, \alpha^{t_j i_2}, \ldots, \alpha^{t_j i_E})$ for $t_j \in T$ and an unknown initial state related vector $X = (c_{i_1} x^{i_1}, c_{i_2} x^{i_2}, \ldots, c_{i_E} x^{i_E})$ such that $v_{t_i} \cdot X = G_t(x) = 0$ for all $t_i \in T$. The rank of the equation system in an "annihilator"-attack has been assumed to have almost full rank E in literature, but it has been an open question. We can now resolve this question by noting that the matrix M is a *generalized* Vandermonde matrix and it was shown by Shparlinski[8] that almost all such matrices have full rank. The *algebraic immunity* of a Boolean function was introduced in [12] and measures the resistance of a function against algebraic attacks. Moreover, the algebraic immunity, abbreviated $AI(f)$, is defined as the minimal degree of a multiple of either f or $f+1$. It has been shown that $AI(f)$ for a k-variable function satisfy the bound $0 \leq AI(f) \leq \lceil k/2 \rceil$. The adversary can therefore always reduce data-complexity from $\sum_{i=0}^{d} \binom{n}{i}$ to roughly $2 \sum_{i=0}^{\lceil \frac{k}{2} \rceil} \binom{n}{i}$ if the degree of the function f is larger than $AI(f)$. But all hope is not lost even if the design employs a Boolean function with optimal algebraic immunity. It was shown in [22], that if there exist polynomials g and h with $\deg(g) < \deg(h) < \deg(f)$ where $h = g \cdot f$, the adversary can instead set up an equation system of the form

$$h_t(S_0) + g_t(S_0) \cdot z_t = 0$$

for $t = 0, 1, 2, \ldots$. Let $e = \deg(g)$, $d = \deg(h), E = \sum_{i=0}^{e} \binom{n}{i}$ and $D = \sum_{i=0}^{d} \binom{n}{i}$. Further, let $H_t(x) = h_t(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x))$ and $G_t(x) = g_t(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x))$ such that

$$H_t(x) + G_t(x) \cdot z_t = 0.$$

The authors of [7] showed that if the adversary pre-computes the minimal polynomial $m_h(x)$ of the sequence $b_t = h(S_t)$ she can simply apply the recursion defined by $m_h(x) = \sum_{i=0}^{D} c_i x^i$ to the equation system

$$\sum_{i=0}^{D} c_i (H_{t+i}(x) + G_{t+i}(x) z_{t+i}) = 0$$

for $t = 0, 1, 2, \ldots, E-1$. The polynomial $m_h(x)$ is simply $\prod_{c_i \neq 0} (x + \alpha^i)$ where c_i are the coefficients of $H_0(x)$ where we assume that all the coefficients for terms x^i of weight less or equal to d are nonzero. Since the sequence $h(S_t) = H_t(x)$ obeys the recursion defined by $m_h(x)$, the new equations become

$$\sum_{i=0}^{D} c_i (G_{t+i}(S_0) z_{t+i}) = 0$$

for $t = 0, 1, 2, \ldots$ which is now a system of equations involving the E coefficients of $G_t(x)$. The best total complexity for solving such equation systems has been shown to be $O(EDlog_2(D) + E^{log_2(7)})$. It is assumed that one needs $D + E$

keystream bits to solve this system, since the relation D is used to determine E equations. However, in practice one can compute a polynomial of degree $D-E$ and zeros α^i with $e < wt(i) \leq d$ that will cancel only the terms of x^i where i has weight larger than e, so only D keystream bits are needed in practice. However, $O(D+E) = O(D)$ for typical applications, so it usually makes little or no difference.

3 Filter Generators in the Spirit of Welch-Gong

The Welch-Gong type filter generator consists of a primitive LFSR over an extension field $\mathbb{F}_q = \mathbb{F}_{2^k}$ of length n and a Boolean function $f(x)$ over $\mathbb{F}_q$. Let $\alpha \in \mathbb{F}_{q^n}$ denote a root of the LFSR generator polynomial. The LFSR defines a q-ary sequence

$$s_t = L(x\alpha^t)$$

where $L(x) = Tr_{q^n/q}(x) = \sum_{i=0}^{n-1} x^{q^i}$ denotes the trace from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$ and $x \in \mathbb{F}_{q^n}$ is a random nonzero initial state. The WG-design applies a Boolean function $f(x)$ to exactly one q-ary element $L(x\alpha^t)$ of the LFSR register, in effect generating a binary keystream

$$z_t = f(L(x\alpha^t))$$

for $t = 0, 1, 2 \ldots$. In the following section our focus will be on minimizing the complexity of univariate algebraic attacks on this particular construction.

3.1 Powers of Subfield Polynomials and Minimum Distance

When solving univariate equations we do not care so much about degree as we care about the number of nonzero coefficients in the polynomials. The equations we are interested in are of the form

$$\begin{aligned} z_t =& f(L(x\alpha^t)) \\ =& \sum_{i=0}^{q-1} c_i L(x\alpha^t)^i \\ =& F(x\alpha^t) \end{aligned}$$

where f is over $\mathbb{F}_q$ and $F(x)$ is over $\mathbb{F}_{q^n}$. In the rest of the paper we will write capital letters F, G, H to represent functions f, g, h over $\mathbb{F}_q$ composed with $L(x)$, where $L(x)$ will be fixed in the context. Notice that we need not take the composition $f(L(x))$ modulo $x^{q^n} + x$ since the highest degree term possible in $L(x)^{q-1}$ is $q^{(n-1)}(q-1) = q^n - q^{n-1}$. To any polynomial $f(x)$ over $\mathbb{F}_q$, define a weight enumerator polynomial

$$T_f(x) = \sum_{i=0}^{k} w_i x^i \tag{2}$$

where w_i counts the number of nonzero terms x^d in f with exponent d of hamming weight i. We have that $T_f(1)$ is the usual hamming weight if the coefficients of f are binary. If $f(x), L(x)$ are as above it is easy to determine the number of nonzero coefficients of their composition $F(x)$.

Theorem 1. *The number of nonzero coefficients of $F(x) = f(L(x))$ is given by $T_f(n)$.*

Proof. In the expansion of $L(x)^e$ with $e = 2^{u_1} + 2^{u_2} + \ldots, 2^{u_d}$ with $u_{i-1} < u_i$ we get terms of the form

$$x^{q^{i_1}2^{u_1}+q^{i_2}2^{u_2}+\cdots q^{i_d}2^{u_d}}. \tag{3}$$

Each varying i and j in the range $0 \leq i < n$ and $0 \leq j < k$, we get that $q^i 2^j$ corresponds to nk distinct powers $2^0, 2^1, \ldots, 2^{nk-1}$. Moreover, since $0 \leq u_1 < u_2 < \ldots < u_d < k$, the exponents in (3) must correspond to distinct integers of hamming-weight d. In a general sum of powers of L,$f(L(x)) = \sum_{i=0}^{2^k-1} L(x)^i$, the terms of weight d (3) can be reached by varying over each possible $\binom{k}{d}$ choice of $u = (u_1, u_2, \ldots, u_d)$ (corresponding to different powers of $L(x)$). Let $A_u = \{q^i 2^u \,|\, 0 \leq i \leq n-1\}$. Notice that $A_u \cap A_{u'} = \emptyset$ for $u \neq u^{'}$. Each distinct choice of u gives rise to n^d exponents

$$A_{u_1} + A_{u_2} + \ldots + A_{u_d} = \{b_1 + b_2 + \ldots + b_d \,|\, b_i \in A_{u_i}\}$$

where, since $u_1 < u_2 < \ldots < u_d$, all must be distinct. Since each choice of $0 \leq u_1 < u_2 < \ldots < u_d$ for a $0 \leq d < k$ gives rise to $|A_{u_1} + A_{u_2} + \ldots + A_{u_d}| = n^d$, it follows that $T_f(n) = \sum_{i=0}^{n} w_i n^i$ is the the number of nonzero coefficients . □

Due to this special structure of $F(x)$ we can improve the bounds on the so-called *spectral immunity* of univariate polynomials. Spectral immunity of a general Boolean polynomial $F(x)$ over $\mathbb{F}_{q^n}$ was essentially defined in [11] in terms of sequences, but here it is more convenient to use the definition provided by Helleseth et. al. [21] in terms of cyclic codes.

Theorem 2. *The spectral immunity of a Boolean function $F(x)$ over $\mathbb{F}_{q^n}$, denoted $SI(F)$, is equal to the minimum weight of a q-ary cyclic code generated by*

$$G_F(x) = \gcd(F(x) + 1, x^{q^n} + x)$$

or

$$G_{F+1}(x) = \gcd(F(x), x^{q^n} + x).$$

The spectral immunity is the univariate analog of algebraic immunity as it measures the least number of unknowns one needs to solve for in an algebraic attack. In a general algebraic attack over $\mathbb{F}_{2^n}$ (when the LFSR is defined over $\mathbb{F}_2$), we have polynomials of the form

$$P(x) = f(L_{t+i_1}(x), L_{t+i_2}(x), \ldots, L_{t+i_k}(x))$$

Since the algebraic immunity of $f(x)$ as a multivariate polynomial in k variables is at most $\lceil k/2 \rceil$, it follows that the spectral immunity of $P(x)$ is upper-bounded by $\sum_{i=0}^{\lceil k/2 \rceil} \binom{n}{i}$. Although univariate and multivariate attacks have similar complexity in general, as we have seen, the WG-type construction produces polynomials of a very special type that allows us to improve this bound significantly.

Lemma 1. *Let $F(x)$ be of the above form (essentially defining a WG-cipher). The minimum distance of the cyclic codes generated by G_F and G_{F+1} over $\mathbb{F}_{q^n}$ is upper-bounded by*

$$SI(F) \leq \sum_{i=0}^{\lceil k/2 \rceil} \binom{k}{i} n^i - (\binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil} - 1) n^{\lceil k/2 \rceil}.$$

Proof. The proof is straight-forward. Assume the worst case, which is when $f(x)$ is balanced. The matrix M_i containing the $\sum_{i=0}^{\lceil k/2 \rceil} \binom{k}{i}$ coefficient vectors of $x^d \pmod{g_i(x)}$ where $g_i(x) = \gcd(f(x) + i, x^q + x)$, $i \in \mathbb{F}_2$ and with d of hamming weight less or equal to $\lceil k/2 \rceil$, has rank at most 2^{k-1}. Consequently, the kernel K_i of M_i has dimension at least $\sum_{i=0}^{\lceil k/2 \rceil} \binom{k}{i} - 2^{k-1} = \binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil}$. Since $\binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil} = \binom{k}{\lceil k/2 \rceil}$ if k is odd and equal to $\binom{k}{\lceil k/2 \rceil}/2$ if k is even, it follows that there exist for each of f and $f+1$ a multiple g with coefficient vector in K_i with at most $\binom{k}{\lceil k/2 \rceil} - \binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil} + 1$ terms x^d where d has hamming weight $\lceil k/2 \rceil$. If $s(x)$ is the polynomial with these coefficients, $T_s(n)$ yields the desired upper-bound. □

It is not clear whether the upper-bound for $T_g(n)$ for a multiple g of f or $f+1$ (the least number of coefficients of a multiple G of F or $F+1$) is tight or not, and leave this as an open problem.

3.2 Minimizing Data Complexity In Attacks on WG Ciphers

In our attacks on the WG-ciphers we seek to minimize the data-complexity. To do this we focus on relations $f \cdot g = h$ where h and g only have terms x^i of weight at most $\lceil k/2 \rceil$. Let M_i denote the matrix spanned by the coefficient vectors of $x^d \pmod{r_i(x)}$ where $r_i(x) = \gcd(f(x) + i, x^q + x)$ for $i \in \mathbb{F}_2$ and where $wt(d) \leq \lceil k/2 \rceil$. When k is odd we have that $\binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil} = \binom{k}{\lceil k/2 \rceil}$, and when k is even it is equal to $\binom{k-1}{\lceil k/2 \rceil} = \binom{k}{\lceil k/2 \rceil}/2$. When k is odd the kernel K_i has at least dimension $\binom{k}{\lceil k/2 \rceil}$, so there must exist multiples $g_0(x)$ of $f(x)$ and $g_1(x)$ of $f(x)+1$ with at most one (and the same) term x^i with exponent weight $\lceil k/2 \rceil$ such that their sum $g(x) = g_0(x) + g_1(x)$ has only terms of weight $< \lceil k/2 \rceil$. In particular, since $f(g_0 + g_1) \equiv g_0 \pmod{x^q + x}$ and $(f+1)(g_0 + g_1) \equiv g_1 \pmod{x^q + x}$ any equation

$$z_t + f(L_t(x)) = 0$$

with f over $\mathbb{F}_{2^k}$ with k odd can always be replaced by an equation

$$0 = g(L_t(x))(z_t + f(L_t(x))) = g(L_t(x))z_t + g_0(L_t(x)) \tag{4}$$

where $G(x) = g(L(x))$ has at most $\sum_{i=0}^{\lceil k/2 \rceil -1} \binom{k}{i} n^i$ nonzero coefficients while $G_0(x)$ has at most $\sum_{i=0}^{\lceil k/2 \rceil -1} \binom{k}{i} n^i + n^{\lceil k/2 \rceil}$ nonzero coefficients.

Similar argument can be made when k is even. Since in this case the kernels have at least dimension $\binom{k-1}{\lceil k/2 \rceil}$, we can find a polynomial g_0 in K_0 and g_1 in K_1 such that the two polynomials share at least $\binom{k-1}{\lceil k/2 \rceil} - 1$ terms and coefficients of weight $\lceil k/2 \rceil$. Since their sum $g(x) = g_0(x) + g_1(x)$ has at most $\binom{k-1}{\lceil k/2 \rceil} + 1$ terms of weight $\lceil k/2 \rceil$, we can construct equations of the form

$$0 = g(L_t(x))(z_t + f(L_t(x))) = g(L_t(x))z_t + g_0(L_t(x))$$

but where $g_0(L_t(x))$ now has at most $\sum_{i=0}^{\lceil k/2 \rceil -1} \binom{k}{i} n^i + (\binom{k-1}{\lceil k/2 \rceil} - 1) n^{\lceil k/2 \rceil}$ nonzero coefficients. Let $E = \sum_{i=0}^{\lceil k/2 \rceil -1} \binom{k}{i} n^i$ and $D = \sum_{i=0}^{\lceil k/2 \rceil} \binom{k}{i} n^i - (\binom{2 \cdot \lceil k/2 \rceil -1}{\lceil k/2 \rceil} - 1) n^{\lceil k/2 \rceil}$ such that the complexity of a fast algebraic attack on a generic WG cipher is upper-bounded by $C = EDlog(D) + E^{log_2(7)}$. Then if k is small in comparison to n, C is roughly equal to

$$O((\binom{k}{\lceil k/2 \rceil - 1} n^{\lceil k/2 \rceil - 1})^{log_2(7)}) \tag{5}$$

and data complexity of $O(n^{\lceil k/2 \rceil})$ keystream bits.

4 Cryptanalaysis of the WG Family

The class of WG-ciphers are pure filter generator constructions consisting of an LFSR of length n over $\mathbb{F}_q = \mathbb{F}_{2^k}$ and a k-variabe Boolean function $f(x)$ over $\mathbb{F}_q$. In this section we analyse WG-5, WG-7, WG-8 and WG-16. In the following let $L(x)$ denote the trace polynomial from $\mathbb{F}_{q^n}$ to $\mathbb{F}_q$ where n and $q = 2^k$ is clear from the context.

4.1 Breaking WG-5

For WG-5 [15] we have $n = 32$ and $\mathbb{F}_q = \mathbb{F}_{2^5}$. The Boolean function is given by $f(x) = \mathrm{Tr}(x^d)$ over $\mathbb{F}_q$ where the specification leaves a choice of either using $d = 7$ or $d = 15$. Since the functions have optimal algebraic immunity, the designers state that the best possible algebraic attack has complexity 2^{54} using 2^{19} keystream bits. By using our rough bounds we can show that there must exist g and h where $g(L(x))$ has $n^2 = 2^{10}$ nonzero coefficients and $h(L(x))$ has $n^3 = 2^{15}$ coefficients. One can verify that $g_0(x) = x^{24} + x^8 + x^7 + x^5 + y$ and $g_1(x) = x^{24} + x^9 + x^8 + x^7 + x^5 + y$ satisfy $f(g_0 + g_1) + g_0 = 0$. Moreover, from $f(x) \cdot g(x) = g_0(x)$, we get equations of the form

$$z_t \cdot L(x\alpha^t)^9 = g_0(L(x\alpha^t))$$

for $t = 0, 1, 2, \ldots$ where the left-hand side involves $n^2 = 2^{10}$ unknowns and the right-hand side roughly $n^3 = 2^{15}$ unknowns. Thus the complexity of a fast algebraic attack is roughly $(n^2)^{\log(7,2)} \approx 2^{30}$ using $n^3 = 2^{15}$ keystream bits. We find $\binom{5}{3} = 10$ such relations that can be used to mount the same attack for both $\mathrm{Tr}(x^7)$ and $\mathrm{Tr}(x^{15})$. Although using the function $\mathrm{Tr}(x^{15})$ result in a higher linear complexity than $\mathrm{Tr}(x^7)$ against an algebraic attack, they behave the same against our attack.

4.2 Breaking WG-7

WG-7 [16] consists of an LFSR of length 23 over a field $\mathbb{F}_{2^7}$ and a Boolean function

$$f(x) = \mathrm{Tr}(x^3 + x^9 + x^{21} + x^{57} + x^{87})$$

corresponding to a multivariate Boolean function in seven variables. The linear complexity of the keystream generated by WG-7 is approximately $2^{25.5}$ and the authors assume that an attacker has access to no more than 2^{24} keystream bits. If the function has optimal algebraic immunity the complexity of an algebraic attack is roughly 2^{69} using 2^{25} keystream bits. Using our rough bounds, the complexity is at most 2^{38} using 2^{18} keystream bits, which is much less than 2^{25}. But the authors of [23] find that the algebraic immunity $f(x)$ is in fact 3, and mount an algebraic attack in 2^{28} using $2^{19.38}$ keystream bits. Due to the low algebraic immunity, it is easy to find exactly one low weight multiple $g_0(x)$ of $f(x)$ and $g_1(x)$ for $f(x)+1$ where all coefficients for both polynomials are 1 and all exponents have hamming weight less or equal to 3. The sum of $g_0(x)$ and $g_1(x)$ is simply $g(x) = g_0(x) + g_1(x) = \mathrm{Tr}(x) = \sum_{i=0}^{7} x^{2^i}$ which can be used to construct a set of equations

$$z_t \, \mathrm{Tr}(L(x\alpha^t)) = g_0(L(x\alpha^t))$$

for $t = 0, 1, 2, \ldots$. The number of unknowns in the right-hand side equation is given by $T_{g_0}(n) = 2^{13.84} \approx 2^{14}$ and the left-hand side has $n \cdot k = 161$ unknowns. Moreover, a FAA on this uses only 2^{14} keystream bits and has complexity about $(n \cdot k) \times 2^{14} \log_2(2^{14}) + (n \cdot k)^{\log(7)} \approx 2^{25}$. This is a factor 2^3 faster than the algebraic attack of [23], but more importantly uses only a factor 2^5 of the their keystream complexity which has the practical significance in this setting.

4.3 Breaking WG-8

WG-8 [17] consists of an LFSR of length 23 over $\mathbb{F}_q = \mathbb{F}_{2^8}$ and apply the Boolean function

$$f(x) = \mathrm{Tr}(x^9 + x^{37} + x^{63} + x^{127})$$

over $\mathbb{F}_q$. The authors claim that the best algebraic attack on this construction is in 2^{69} using 2^{26} keystream bits. To find good relations for this specification, we computed the kernel of M_0 and M_1 consisting of rows spanned by the coefficients of $x^t \pmod{g_i(x)}$ for all t of weight ≤ 4 and where $g_i(x) = \gcd(f(x)+i, x^q+x)$.

Each of the kernels had minimal dimensions 35, and their sum $K = K_0 + K_1$ dimension 70. We simply row reduced the basis matrix for K to eliminate the high weight terms first and collected the last row of the matrix. It contained the coefficient vector of a polynomial $s(x) = s_0(x) + s_1(x)$ with only one term x^d of weight 4 and had $E = T_s(n) = 2^{17}$. Moreover, the polynomials $s_i(L(x))$ have $D = T_{s_i}(n) \approx 2^{22}$ nonzero coefficients. In an attack we can therefore use a relation

$$z_t s(L(x\alpha^t)) + s_0(L(x)) = 0$$

for $t = 0, 1, 2, \ldots$ and obtain an attack with complexity $EDlog(D)+E^{log(7)} \approx 2^{48}$ using 2^{22} keystream bits. We found many relations that gives the same attack complexity.

4.4 Breaking WG-16

WG-16[18] consists of a primitive LFSR of length 32 over $\mathbb{F}_q = \mathbb{F}_{2^{16}}$ and is meant for use in 4G. The function $f(x)$ has multivariate degree 8 and optimal algebraic immunity. The authors claim that the best algebraic attack on this construction is in 2^{159} using 2^{58} keystream bits. Since the function is in an even number of variables, the minimal $T_g(n)$ for a multiple g of f and $f + 1$ satsify $T_g(n) \leq \sum_{i=0}^{\lceil k/2 \rceil -1} \binom{k}{i} n^i + (\binom{k-1}{\lceil k/2 \rceil} + 1) n^{\lceil k/2 \rceil} \approx 2^{53}$. A direct univariate algebraic attack has then complexity $T_g(n)^{\log(7,2)} = 2^{148}$ using 2^{54} keystream bits which is already less than the claimed bounds. Using the bounds for relations $f \cdot g = h$, there must exist a polynomials g and h where $g(x)$ has exactly one term x^i with $wt(i) >= \lceil k/2 \rceil - 2$ and $h(x)$ has terms of weight up to $\lceil k/2 \rceil + 2$. Thus we can set $E = T_g(n) \leq \sum_{i=0}^{\lceil k/2 \rceil -2} \binom{k}{i} n^i + n^{\lceil k/2 \rceil} \approx 2^{37}$ and $D = T_h(n) \leq \sum_{i=0}^{\lceil k/2 \rceil +2} \binom{k}{i} n^i \approx 2^{63}$ that yields an attack with computational complexity $ED \log(D) + E^{\log(7,2)} \approx 2^{106}$ using 2^{63} keystream bits which, assuming that the key has size 128, breaks the cipher.

5 Conclusion

In this paper we have described practical applications of certain cyclic codes over $\mathbb{F}_q$ and $\mathbb{F}_{q^n}$ generated by a Boolean function. Determining the immunity against an algebraic attack on a WG-type construction involves determining the minimum value $T_s(n)$ where s are codewords of the cyclic codes generated by $g_0(x) = \gcd(f(x), x^q + x)$ and $g_1(x) = \gcd(f(x) + 1, x^q + x)$ over $\mathbb{F}_q$. This is a slightly different problem than finding minimum distances of a code and an algorithm for determining this *ordered minimum distance* problem is left as an open problem. Moreover, it does not seem that maximal algebraic immunity alone is sufficient to ensure optimal values for $T_s(n)$. We propose that a strong Boolean function should attain maximal value among the codewords s in its two cyclic codes and that the analysis of this paper must be accounted for when designing secure word-based stream ciphers.

It should also be noted that our analysis may have applications to similar word-based constructions, most notably SNOW-3G [13].

References

1. R. Lidl and H. Niederreiter, Finite Fields *In Encyclopedia of Mathematics and its Applications, Cambridge University Press, 1997.*
2. *S. W. Golomb, G. Gong, Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar (2004), Cambridge University Press, New York, USA.*
3. *F. Armknecht and M. Krause, Algebraic attacks on combiners with memory,* Advances in Cryptology-CRYPTO 2003*, Lecture Notes in Computer Science, vol. 2729, pp. 162-176, Springer-Verlag, 2003.*
4. *F. Armknecht, Improving fast algebraic attacks,* Proceedings of Fast Software Encryption 2004*, Lecture Notes in Computer Science, vol. 3017, pp. 65-82, Springer-Verlag, 2004.*
5. *Philip Hawkes and Gregory G. Rose., Rewriting variables: The complexity of fast algebraic attacks on stream ciphers.* In Matt Franklin, editor, Advances in Cryptology - CRYPTO 2004: 24th Annual International Cryptology Conference, Santa Barbara, California, USA, August 15-19, 2004. Proceedings, Lecture Notes in Computer Science, Berlin / Heidelberg, 2004. Springer-Verlag.
6. F. Armknecht and G. Ars, Introducing a new variant of fast algebraic attacks and minimizing their successive data complexity, Mycrypt 2005 (International Conference on Cryptology in Malaysia), Lecture Notes in Computer Science, vol. 3715, pp. 16-32, 2005, E. Dawson and S. Vaudenay (Eds.)
7. N. Courtois, Fast algebraic attacks on stream ciphers with linear feedback, *Advances in Cryptology-Crypto'2003*, Lecture Notes in Computer Science, vol. 2729, pp. 176-194, Springer-Verlag, 2003.
8. Shparlinski, Igor E.,On the Singularity of Generalised Vandermonde Matrices over Finite Fields, *In Finite Fields and Applications, vol. 11, no. 2,pp. 193–199, 2005, Elsevier Science Publishers B. V.*
9. *N. Courtois and W. Meier, Algebraic attacks on stream ciphers with linear feedback,* Advances in Cryptology-Eurocrypt'2003*, Lecture Notes in Computer Science, vol. 2656, pp. 345-359, Springer, 2003.*
10. *S.W. Golomb,* Shift Register Sequences, *Holden-Day, Inc., San Francisco, 1967, revised edition, Aegean Park Press, Laguna Hills, CA, (1982).*
11. *G. Gong, S. Rønjom, T. Helleseth and H. Hu, Fast linear subspace attacks on stream ciphers, submitted to* IEEE Transactions on Information Theory.
12. *W. Meier, E. Pasalic, and C. Carlet. Algebraic attacks and decomposition of Boolean functions.* In Advances in Cryptology — EUROCRYPT 2004, volume 3027 of Lecture Notes in Computer Science, pages 474–491. Christian Cachin and Jan Camenisch, editors, Springer, 2004.
13. ETSI/SAGE. Specification of the 3GPP Confidentiality and Integrity Algorithms UEA& UIA2 Document 2: Snow 3G Specification (version 1.1) (September 2006), http://www.3gpp.org/ftp
14. S. Rønjom and T. Helleseth, A New Attack on the Filter Generator, *IEEE Transactions on Information Theory*, vol. 53, no. 5, pp. 17520-1758, 2007.
15. Aagaard, M.D. and Guang Gong and Mota, R.K. Hardware implementations of the WG-5 cipher for passive RFID tags, *In IEEE International Symposium on Hardware-Oriented Security and Trust (HOST), 2013 , June 2013, pp. 29-34.*
16. *Yiyuan Luo and Qi Chai and Guang Gong and Xuejia Lai, A Lightweight Stream Cipher WG-7 for RFID Encryption and Authentication,* In IEEE Global Telecommunications Conference (GLOBECOM 2010), 2010, Dec 2010, pp. 1-6.

17. Fan, Xinxin and Mandal, Kalikinkar and Gong, Guang, WG-8: A Lightweight Stream Cipher for Resource-Constrained Smart Devices, *In Quality, Reliability, Security and Robustness in Heterogeneous Networks,Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering,ed. Singh, Karan and Awasthi, AmitK,Springer Berlin Heidelberg*
18. *Xinxin Fan and Guang Gong, Specification of the Stream Cipher WG-16 Based Confidentiality and Integrity Algorithms,* Technical Report CACR 2013-06 at University of Waterloo, CA: http://cacr.uwaterloo.ca/techreports/2013/cacr2013-06.pdf
19. Yassir Nawaz , Guang Gong, The WG Stream Cipher. *EU ECRYPT eSTREAM competition,http://www.ecrypt.eu.org/stream/*
20. *S. Rønjom and T. Helleseth, Attacking the filter generator over* $GF(2^m)$*,* Arithmetic of Finite Fields, First International Workshop, WAIFA 2007, Madrid, Spain, June 2007, Lecture Notes in Computer Science*, vol. 4547, pp. 264-275, 2007.*
21. *T. Helleseth and S. Rønjom. Simplifying algebraic attacks with univariate analysis.* In Information Theory and Applications Workshop (ITA), 2011, pages 1–7, Feb. 2011.
22. N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. *In Advances in Cryptology - CRYPTO 2003, number 2729 in Lecture Notes in Computer Science, pages 176–194. Springer Verlag, 2003.*
23. *Orumiehchiha, MohammadAli and Pieprzyk, Josef and Steinfeld, Ron, Cryptanalysis of WG-7: a lightweight stream cipher,* In Cryptography and Communications,volumne 4, nr. 3-4, 2012, Springer US.