Improving algebraic attacks on stream ciphers based on linear feedback shifter registers over $\mathbb{F}_{2^k}$

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 reduce the general algebraic attack complexity on such cipher by proving new and lower bounds for the spectral immunity of such ciphers. 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. In particular, there is an algebraic degeneracy in these constructions, which, when combined with attacks based on low-weight multiples over $\mathbb{F}_q$, provides much more efficient attacks than over $\mathbb{F}_2$. With negligible computational complexity, our best attack breaks the primitive WG-5 if given access to 4 kilobytes of keystream, break WG-7 if given access to 16 kilobytes of keystream and break WG-8 if given access to half a megabyte of keystream. Our best attack on WG-16 targeted at 4G-LTE is less practical, and requires 2^{103} computational complexity and 2^{61} bits of keystream. In all instances, we significantly lower both keystream and computational complexity in comparison to previous estimates. On a side note, we resolve an open problem regarding the rank of a type equation systems used in algebraic attacks.

Keywords: Cryptography, spectral immunity, 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]. All of the constructions are based on generating a keystream sequence based on filtering a single word from a shift register over a finite field of size 2^k, where k is indicated in the name of the cipher (e.g. WG-k means the shift-register is over $\mathbb{F}_{2^k}$). We show that the general complexity for an attack over the extension field (depending only on the size n of the LFSR and the size k of the field) is typically less or much less than multivariate attacks over $\mathbb{F}_2$. These designs are rich in algebraic structure when viewed over the extension field in terms of univariate polynomials, which has already been utilized in previous papers in [20] and [21]. Moreover, by combining a

theorem by Brynielsson [23] together with a the spectral immunity as described in Helleseth et al in [14], we are able to significantly reduce the spectral immunity and complexity of algebraic attacks on WG-type constructions.

Cipher	Bit security	Computational	Data	Ref
WG-5 (7/15)	80	2^{51}	$2^{19.3}$	[15]
WG-5 (7/15)	80	2^{43}	2^{15}	New (bound SI)
WG-5 (7/15)	80	2^{33}	2^{18}	New (bound SI)
WG-5 (7)	80	2^{31}	2^{15}	New (found)
WG-5 (15)	80	2^{33}	2^{15}	New (found)
WG-7	80	$2^{66.1}$	$2^{24.7}$	[16]
WG-7	80	2^{30}	$2^{19.3}$	[22]
WG-7	80	$2^{54.6}$	$2^{19.5}$	New (bound SI)
WG-7	80	2^{43}	$2^{23.3}$	New (bound)
WG-7	80	2^{28}	$2^{17.3}$	New (found)
WG-8	80	2^{69}	2^{26}	[17]
WG-8	80	$2^{63.4}$	$2^{22.6}$	New (bound SI)
WG-8	80	2^{54}	$2^{23.5}$	New (bound)
WG-8	80	2^{48}	2^{22}	New (found)
WG-16	128	2^{159}	2^{58}	[18]
WG-16	128	2^{148}	$2^{52.7}$	New (bound SI)
WG-16	128	$2^{106.5}$	2^{63}	New (bound)
WG-16	128	2^{102}	2^{61}	New (found)

Table 1. Summary of bounds versus found algebraic attack complexity on the WG-ciphers. SI=Spectral Immunity

Table 1 above summarizes our algebraic attacks on the WG-ciphers. For WG-5, WG-7 and WG-8, the computational complexity of an attack is negligible. Thus, on these ciphers the focus is on reducing data-complexity, which is the dominant factor when considering real world applications.

We are not aware of previously published attacks against the 80-bit key stream cipher WG-5. The designers provide two versions and estimate that the best algebraic attacks on these two requires a data complexity of $2^{19.35}$ bits and 2^{51} computational complexity. Thus, this design is already moving along the edge of security. Our contribution is to show that both versions can be attacked using 2^{15} keystream bits and with negligible computational complexity.

The 80-bit cipher WG-7 has already been broken with an algebraic attack over $\mathbb{F}_2$ in [22], using $2^{19.38}$ keystream bits and negligible time complexity. Their attack use the fact that the Boolean function only has algebraic immunity 3. In addition to reducing the keystream complexity to 2^{17}, we show that even if

the Boolean function had optimal algebraic immunity 4, a univariate algebraic attack would still use roughly 2^{19} bits of keystream.

The best attack on the 80-bit cipher WG-8 is a related IV attack using 2^{52} chosen IVs and computational complexity $2^{53.32}$([24]). The best algebraic attack on this cipher is estimated to be 2^{66} computational complexity using $2^{24.65}$ keystream bits. In this paper, we provide a key-recovery attack on WG-8 using 2^{22} keystream bits and with computational complexity of 2^{48}.

The best attack on the 128-bit cipher WG-16 is estimated to be 2^{156} computational complexity using 2^{57} keystream bits. We have found relations that reduce this to 2^{102} computational complexity using 2^{61} keystream bits.

Previous papers on algebraic cryptanalysis of stream ciphers based on linear shift-registers have mainly focused on solving equations over $\mathbb{F}_2$. However, it has recently been shown (see for instance [21]) that univariate algebraic attacks over extension fields can be more efficient, as the rich structures of the underlying finite fields become more visible. Thus, representation can make a big difference in the case of stream ciphers based on linear feedback shift registers.

Our main contribution is to show that the minimal keystream requirement in an algebraic attack on certain filter generators based on LFSRs over extension fields are generally much less than previously known or estimated. This complements and improves a result in [20]. We show this by applying a theorem of Brynielsson to reduce the bounds on the spectral immunity in our cryptanalysis that significantly reduces the complexity in algebraic attacks on such. A direct application of our results yields new and lower bounds on the keystream complexity in such attacks. For WG-5, WG-7, WG-8 and WG-16 we have confirmed that the complexity is actually less than our new bound.

On a side-note, we solve a long-standing open problem in literature regarding the rank of the equation systems derived from using so-called annihilator equations (see [25]). We identify that the coefficients in these equation systems span generalized Vandermonde matrices, for which it has been shown by Spharlinski has almost always full rank. Though this has long been believed to be true, it has not before been confirmed.

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 initialized 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 initialized 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 consist 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 exists 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, but it has remained an open question in literature ([25]). 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 satisfies 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 [**?**], 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 The spectral immunity of WG-type stream ciphers

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, the number of nonzero coefficients in their composition follows directly from a theorem by Brynielsson.

Theorem 1 ([23]). *The number of nonzero coefficients of $F(x) = f(L(x))$ is given by $T_f(n)$.*

Due to this special structure of $F(x)$ we can now improve the bounds on the *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 polynomials and cyclic codes.

Theorem 2 ([21]). *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, thus minimizing the data complexity. 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.

Theorem 3. *Let $F(x) = f \circ L(x)$ where $f(x)$ is defined over $\mathbb{F}_{2^k}$ and $L(x)$ is a trace from $\mathbb{F}_{2^{nk}}$ to $\mathbb{F}_{2^k}$. The minimum distance of the cyclic codes generated by G_F and G_{F+1} over $\mathbb{F}_{2^{nk}}$ 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 $\sum_{r=0}^{\lceil k/2 \rceil} \binom{k}{r}$ 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 $\sum_{r=0}^{\lceil k/2 \rceil} \binom{k}{r} - 2^{k-1} = \binom{2 \cdot \lceil k/2 \rceil - 1}{\lceil k/2 \rceil}$. 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. □

Notice that $SI(F) \leq \sum_{i=0}^{\lceil k/2 \rceil - 1} \binom{k}{i} n^i + n^{\lceil k/2 \rceil}$ when k is odd and $SI(F) \leq \sum_{i=0}^{\lceil k/2 \rceil - 1} \binom{k}{i} n^i + \binom{k-1}{\lceil k/2 \rceil} n^{\lceil k/2 \rceil}$ when k is even. As a consequence, the spectral immunity of WG-ciphers is much worse in comparison to equivalently sized binary filter generators consisting of a LFSR of size $n \cdot k$ and a Boolean function in k variables that tap bits from k arbitrary positions of the LFSR. In particular, both the keystream complexity and the computational complexity of an algebraic attack is in general lower than previously thought.

Note that 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 interesting open problem.

3.2 Minimizing Complexity In Attacks on WG Ciphers

In the following we shed some light on the consequences of Theorem 3. The proof of Theorem 3 instructs us how to find other types of relations that do not minimize the keystream complexity. By adjusting Theorem 3 on the minimal $T_{g_i}(n)$ (spectral immunity) when using multiples

$$(f(x)+i)g_i(x) = g_i(x),$$

we can derive bounds for $T_g(n)$ and $T_{g_i}(n)$ when using relations of the form

$$(f(x)+i)g(x) = g_i(x)$$

where we allow $g_i(x)$ to have terms with exponents up to and including weight $d \geq \lceil k/2 \rceil$. The following theorem gives an especially simple relationship between the values of $T_{g_i}(n)$ and $T_g(n)$ relative to d.

Theorem 4. *Let $F(x) = f \circ L(x)$ where $f(x)$ is defined over $\mathbb{F}_{2^k}$ and $L(x)$ is a trace from $\mathbb{F}_{2^{nk}}$ to $\mathbb{F}_{2^k}$. For $\lceil k/2 \rceil \leq d \leq k$, there exist $g(x)$ and $g_i(x)$ such that*

$$(f(x)+i)g(x) = g_i(x)$$

where

$$T_{g_i}(n) \leq \sum_{i=0}^{d} \binom{k}{i} n^i$$

and

$$T_g(n) \leq \sum_{i=0}^{k-d-1} \binom{k}{i} n^i + n^{k-d}.$$

Proof. In this case, we compute matrices M_i formed by the coefficients vectors of $x^t \mod \gcd(f(x)+i, x^{2^k}+x)$ where the exponent can have weight $d > \lceil k/2 \rceil$. Assuming that the function $f(x)$ is optimal, the kernel K_i of M_i has dimension $\dim(K_i) = \sum_{r=0}^{d} \binom{k}{r} - 2^{k-1}$. Let $d = \lceil k/2 \rceil + t$. If k is odd, we have that $\sum_{r=0}^{\lceil k/2 \rceil} \binom{k}{r} - 2^{k-1} = \binom{k}{\lceil k/2 \rceil}$, while if k is even, $\sum_{r=0}^{\lceil k/2 \rceil} \binom{k}{r} - 2^{k-1} = \binom{k}{\lceil k/2 \rceil}/2$ Thus in the case of k odd, $\dim(K_0 \oplus K_1) = 2(\binom{k}{\lceil k/2 \rceil} + \sum_{r=\lceil k/2 \rceil+1}^{d} \binom{k}{r})$ is equal to $\sum_{r=\lceil k/2 \rceil - t - 1}^{d} \binom{k}{r}$. Since this is exactly the number of terms x^e with weight between $\lceil k/2 \rceil - t - 1$ and d, it follows that there must exist at least one vector in K that is the coefficient vector of a polynomial $g(x)$ with only one term with exponent weight $\lceil k/2 \rceil - t - 1$. In this case we have that $T_g(n) \leq \sum_{r=0}^{\lceil k/2 \rceil - t - 2} \binom{k}{r} + n^{\lceil k/2 \rceil - t - 1}$ which is equal to $\sum_{r=0}^{k-d-1} \binom{k}{r} n^r + n^{k-d}$, since $2\lceil k/2 \rceil - t - 1 = k - d$ when k is odd.

In the case of k even, $\dim(K_0 + K_1) = 2(\binom{k}{\lceil k/2 \rceil}/2 + \sum_{i=\lceil k/2 \rceil+1}^{d} \binom{k}{i})$ which again is equal to $\sum_{r=\lceil k/2 \rceil - t}^{\lceil k/2 \rceil + t} \binom{k}{r}$. This is exactly the number of terms of weight between $\lceil k/2 \rceil - t$ and d. Moreover, there must exist at least one vector in K that is the coefficient vector of a polynomial $g(x)$ with only one term with exponent weight $\lceil k/2 \rceil - t$. In this case we have that $T_g(n) \leq \sum_{r=0}^{\lceil k/2 \rceil - t - 1} \binom{k}{r} + n^{\lceil k/2 \rceil - t} = \sum_{r=0}^{k-d-1} \binom{k}{r} n^r + n^{k-d}$, since $2\lceil k/2 \rceil - t = n - d$ when k is even. In both cases, $T_{g_i}(n) \leq \sum_{r=0}^{d} \binom{k}{r} n^r$, since the coefficient vector for $g(x)$ is a sum of arbitrary vectors from K_0 and K_1. □

The attacker can thus, if allowed, lower the total computational complexity by increasing the keystream complexity. The above bounds can then be used to measure the resistance against algebraic attacks when increasing the keystream complexity.

4 Cryptanalysis 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-variable Boolean function $f(x)$ over $\mathbb{F}_q$. In this section we analyze WG-5, WG-7, WG-8 and WG-16. In the following let $Tr(x)$ denote the trace polynomial from $\mathbb{F}_q$ to $\mathbb{F}_2$ where $q = 2^k$ is clear from the context.

For each of the WG-ciphers we have provided a table with bounds on the complexities of doing an algebraic attack by using

1) Theorem 3 to show the minimal keystream complexity of an attack, and,
2) Theorem 4 to show the minimal computational complexity of an attack.

We have then compared this to the relations found by computer search.

4.1 On computing relations that minimizes target complexity

We compute matrices M_0 and M_1 whose kernels K_0 and K_1 span the coefficient vectors of $x^t \mod \gcd(f(x)+i, x^{2^k}+x)$ with terms with exponents t of weight less or equal to $\lceil \frac{k}{2} \rceil \leq d < k$. Thus, for equations of the form

$$g_i(x)f(x) = g_i(x)$$

we look for candidates in the kernel K_i with $d = \lceil \frac{k}{2} \rceil$. To determine equations of the form

$$g(x)(f(x)+i) = g_i(x)$$

we look for candidates among the sum $K = K_0 \oplus K_1$. Since K has greater dimension (typically twice) than each of the kernels K_i, we can find $g(x) = g_0(x) + g_1(x)$ with fewer coefficients, but at the possible expense of increasing the number of coefficients in each $g_i(x)$. In some cases, it might even be better to look for $g(x)$ with exponents of higher weight than $\lceil k/2 \rceil$ to minimize the total complexity. This way, we get larger kernels K_i that can be used to further reduce the complexity $T_g(n)$ at the expense of increasing $T_{g_i}(n)$. There is some freedom in choosing which way to optimize, and should be considered case by case.

In the following we describe our results for each of the WG-ciphers. We have implemented the computations for each of the WG-ciphers in C++ using NTL. The source code is made available at request. For each cipher, we provide a table with the best bounds for the complexity of algebraic attacks. In the tables we let $D = T_{g_i}(n)$ and $E = T_g(n)$. The first row describes the minimal complexity of using single multiples of $f(x)+i$, derived from Theorem 3 on the spectral immunity, while the rest of the rows are derived using the bound in Theorem 4 where d is the maximal weight of the exponents of terms allowed to appear in $g_i(x)$.

4.2 Cryptanalysis of WG-5

d	$\log_2(D)$	$\log_2(E)$	$\log_2(E^{\log_2(7)} + ED\log_2(D))$
3	0	15.4	43.2
3	18.4	10.2	32.9
4	22.4	5.0	31.9
5	25.2	0.0	29.9

Table 2. Summary of bounds for algebraic attacks on WG-5.

WG-5 [15] is a stream cipher with a target security of 80-bits. The design specification states that the best algebraic attack has complexity $2^{51.13}$ and using $2^{19.3}$ data complexity and no other attacks have previously been published in

literature. For WG-5 we have parameters $n = 32$ 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 $d = 15$. The two filter functions are

$$f_7(x) = x^{28} + x^{25} + x^{19} + x^{14} + x^7$$

and

$$f_{15}(x) = x^{30} + x^{29} + x^{27} + x^{23} + x^{15}.$$

Table 2 shows that the spectral immunity is at most $2^{15.4}$ and is also the upper-bound on the minimal keystream complexity of an algebraic attack. Thus, there exists an algebraic attack using at most $2^{15.4}$ (roughly 0.005 megabytes) keystream bits and computational complexity roughly 2^{43} for this parameter setting, regardless of which filter function is used. This already violates the designers security estimates.

For $f_7(x)$ we found polynomials $g_0(x)$ and $g_1(x)$ that can be used to mount attacks requiring less keystream than the general bound. For instance, it can easily be verified that

$$g_0(x) = x + x^2 + x^6 + x^{16} + x^{10} + x^{18} + x^{20} + x^{24} + x^{26}$$

is a multiple of $f_7(x)$ with $T_{g_0}(n) = 2^{15.21}$, while

$$g_1(x) = 1 + x + x^2 + x^4 + x^{16} + x^3 + x^6 + x^{12} + x^{18} + x^{20} + x^{24} + x^{28}$$

is a multiple of $f_7(x) + 1$ modulo $x^{32} + x$ with $T_{g_1}(n) = 2^{15.25}$.The function f_{15} behaves worse than $f_7(x)$. For instance, one can verify that

$$g_0(x) = x^{24} + x^8 + x^7 + x^5 + x$$

and

$$g_1(x) = x^{24} + x^9 + x^8 + x^7 + x^5 + x$$

is such that $T_{g_0}(32) = T_{g_1}(32) = 2^{15.1}$ and $g_0(f_{15}(x) + 1) = g_1(x)f_{15}(x) = 0$. Thus, the minimal keystream requirement for attacking WG-5 is roughly 2^{15}. It is of interest to see if we are able to reduce the total complexity without increasing the keystream requirement too much. In Table 2 we see that there exist an attack in 2^{33} using $2^{18.4}$ keystream bits. Thus, we search for multiples $g_0(x)$ and $g_1(x)$ such that $g(x) = g_0(x) + g_1(x)$ results in $T_g(n) < 2^{15}$, while at the same time requiring that $T_{g_i}(n)$ is not much more than 2^{15}. For $f_7(x)$ one can verify that

$$g(x) = 1 + x^4 + x^5 + x^6 + x^{10} + x^{16} + x^{17}$$

and

$$g_1(x) = x^{28} + x^{24} + x^{20} + x^{18} + x^{16} + x^{12} + x^6 + x^4 + x^3 + x^2 + x + 1$$

satisfy

$$(f_7(x)+1)g(x) = g_1(x)$$

with corresponding complexities $T_{g_1}(32) = 2^{15.2}$ and $T_h(32) = 2^{12}$, which is worse than our bound. In the case of f_{15}, we get even better results. One can for instance verify that

$$g(x) = x^5$$

and

$$g_0(x) = x^{28} + x^{20} + x^4 + x^3 + x$$

satisfy $f_{15}(x)g(x) = g_0(x)$ where $T_g(n) = 2^{10}$ and $T_{g_0}(n) = 2^{15}$. 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 2^{30} using 2^{15} keystream bits.

4.3 Cryptanalysis of WG-7

d	$\log_2(D)$	$\log_2(E)$	$\log_2(E^{\log_2(7)} + ED\log_2(D))$
4	0	19.5	54.6
4	23.3	14.5	42.8
5	27.1	9.4	41.3
6	30.1	4.6	39.6
7	32.1	0.0	37.1

Table 3. Summary of bounds for algebraic attacks on WG-7.

WG-7 [16] is a stream cipher with 80-bits security target. This cipher has already been attacked in [22] where the low algebraic immunity was used to mount an attack in 2^{28} using $2^{19.38}$ keystream bits. The cipher 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. While the cryptanalysis in [22] used the fact that the Boolean function has algebraic immunity 3 instead of 4. Thus, this cipher has already been broken. In Table 3, we see that the minimal keystream complexity (spectral immunity) is already 2^{19}, regardless of the algebraic immunity of the function $f(x)$, and a direct algebraic attack is possible with negligible computational complexity if this amount of keystream is available.

However since $f(x)$ has algebraic immunity 3, it is natural to restrict to multiples with exponents of weight less and equal to 3. In fact, if we let $g(x) = \mathrm{Tr}(x) + 1$ over $\mathbb{F}_q$ it follows that

$$f(x)g(x) = g_0(x)$$

is such that $T_{g_0}(23) = 2^{17.4}$ while $T_g(23) = 161$. The keystream complexity is a little less (1/4th) than the attack in [22], while the computational complexity is negligible.

4.4 Cryptanalysis of WG-8

d	$\log_2(D)$	$\log_2(E)$	$\log_2(E^{\log_2(7)} + ED\log_2(D))$
4	0	22.6	63.4
4	23.5	19.2	54.0
5	27.5	14.2	46.5
6	30.9	9.1	45.0
7	33.5	4.4	43.0

Table 4. Summary of bounds for algebraic attacks on WG-8.

WG-8 [17] is a stream cipher targeting 80-bit security. The best algebraic attack on this construction is estimated at 2^{69} computational complexity using 2^{26} bits of keystream. The cipher consists of an LFSR of length 20 over $\mathbb{F}_q = \mathbb{F}_{2^8}$ and applies the Boolean function

$$f(x) = \mathrm{Tr}(x^9 + x^{37} + x^{53} + x^{63} + x^{127})$$

over $\mathbb{F}_q$. In Table 4, we see that the spectral immunity is less or equal to 2^{23}. However, we find a multiple $g_1(x) = \mathrm{Tr}(x^{15}) + \mathrm{Tr}(x^{45}) + 1$ of $f(x)+1$, thus the minimal keystream requirement to attack this cipher is actually $T_{g_0} = 16(n^4) = 2^{21.3}$ using little less than 2^{60} computational complexity.

Moreover, we find a relation $f(x)g(x) = g_0(x)$ with $T_{g_0}(20) = 2^{22.3}$ and $T_g(20) = 2^{17}$. The coefficients of $g(x)$ are $C_g = \{1, 2, 3, 6, 8, 12, 16, 18, 19, 20, 21, 22, 26, 32, 33, 34, 35, 36, 40, 48, 49, 50, 52, 56, 65, 66, 67, 81, 96, 97, 128, 130, 131, 136, 138, 161, 168, 192\}$. Thus, with a data complexity of $2^{22.3}$, an algebraic attack has complexity roughly $(2^{17})^{\log_2(7)} + 2^{22.3} \cdot (22.3) \approx 2^{48}$.

4.5 Cryptanalysis of WG-16

WG-16[18] is a 128-bit WG-cipher that 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

d	$\log_2(D)$	$\log_2(E)$	$\log_2(E^{\log_2(7)} + ED\log_2(D))$
8	0	52.7	148.0
8	53.7	48.5	136.2
9	58.5	43.0	120.7
10	63.0	37.1	106.5
11	67.2	30.9	104.1
12	70.9	24.2	101.3
13	74.3	17.3	97.7
14	77.1	10.6	94.0
15	79.3	5.0	90.7

Table 5. Summary of bounds algebraic attack complexity on WG-16

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 theoretical optimal minimal value for $T_{g_i}(32)$ for a multiple $g_i(x)$ of f or $f+1$ satisfies $T_{g_i}(32) \approx 2^{53}$. Thus a algebraic attack has then complexity $T_g(n)^{\log(7,2)} = 2^{148}$ using minimal amount of 2^{53} keystream bits. If the function is optimal, there should be no attacks using fewer keystream bits. However, we have by computer search found multiples for both $f(x)$ and $f(x)+1$ with $T_{g_i}(32) \approx 2^{51}$. Moreover, in Table 5 we see that there is an attack in 2^{106} that uses 2^{63} bits of keystream. For instance, using $d = 10$, there must exist an attack in at most 2^{106} that uses at most 2^{63} bits of keystream. We have confirmed this has found g and g_0 with $f \cdot g = g_0$ where $T_{g_0}(32) = 2^{61}$ while $T_g(32) = 2^{36}$, thus improving the attack in comparison to the general bound to roughly 2^{102} computational complexity using 2^{61} keystream bits.

5 Conclusion

In this paper we have derived new bounds for the complexity of algebraic attacks on word-based filter generators. In particular, we show that the complexity of an algebraic attack on filter generators formed by filtering a single word from an LFSR over an extension field is typically much less than previously thought. The main reason for this is that the spectral immunity of these filter generator polynomials is much lower than for equivalently sized classical binary filter generator polynomials. In particular, we give new and improved upper-bounds for the spectral immunity for this class of stream ciphers. The main new result is derived from a theorem by Bryniellson and a theorem by Helleseth et al. on the spectral immunity. As an application, we improve both computational complexity and keystream complexity in algebraic attacks on all the WG-ciphers. A consequence is that designers of stream ciphers similar to these, must carefully evaluate security using Theorem 3 and Theorem 4.

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 Discrete Fourier Spectra Attacks on Stream Ciphers,* In Information Theory, IEEE Transactions on , vol.57, no.8, pp.5555-5565, Aug. 2011
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. *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.
23. L. Brynielsson, On the linear complexity of combined shift register sequences, *EUROCRYPT85*, H.C William (Eds), pp.156-160, Springer, 1985.
24. L. Ding, C. Jin, Cryptanalysis of lightweight WG-8 stream cipher,IEEE Transactions on Information Forensics and Security, Vol. 9, Issue4, 2014, pp. 645-652.
25. A. Canteaut, Open problems related to algebraic attacks on stream ciphers, Coding and Cryptography, LNCS 3969,pp. 120-134, Springe,r 2006.