Resilience to Distinguishing Attacks on WG-7 Cipher and Their Generalizations

Guang Gong, Mark Aagaard and Xinxin Fan
Department of Electrical and Computer Engineering
University of Waterloo, Waterloo, Ontario, Canada
Emails: {ggong, maagaard, x5fan}@uwaterloo.ca

Abstract

The stream cipher WG-7 is a lightweight variant of the well-known Welch-Gong (WG) stream cipher family, targeting for resource-constrained devices like RFID tags, smart cards, and wireless sensor nodes. Recently, a distinguishing attack was discovered against the stream cipher WG-7 by Orumiehchiha, Pieprzyk and Steinfeld. In this paper, we extend their work to a general distinguishing attack and suggest criteria to protect the WG stream cipher family from this attack. Our analysis shows that by properly choosing the minimal polynomial of the linear feedback shift register for a WG stream cipher, the general distinguishing attack can be easily thwarted.

1 Introduction

The Welch-Gong (WG) stream cipher family [6] is a set of synchronous stream ciphers submitted to the ECRYPT Stream Cipher (eSTREAM) Project in 2005. Among more than 20 submissions, the WG stream cipher family is the only candidate that has mathematically proven randomness properties such as ideal two-level autocorrelation, long period, ideal tuple distribution, and exact linear complexity. Those properties are paramount for protecting communication systems from various malicious attacks by attackers and increase the robustness of the signal transmission through noisy wireless communication networks. The WG stream ciphers are hardware-oriented stream ciphers that use a word-oriented linear feedback shift register (LFSR) and a filter function based on the WG transformation [2]. Depending on application scenarios, the WG stream cipher family can be parameterized to provide security solutions for a wide range of embedded applications such as smart cards, wireless sensor nodes,

mobile phones, etc. For example, the stream cipher WG-7 [5] is a lightweight variant of the WG stream cipher family tailored for securing low-cost RFID tags.

Recently, Orumiehchiha *et al.* [7] proposed a distinguishing attack on the stream cipher WG-7. The authors built two distinguishers and showed that the key stream generated by WG-7 can be distinguished from a random sequence after obtaining $2^{13.5}$ keystream bits and with a nonnegligible probability. While the proposed distinguishing attack does not affect the security of the WG stream cipher in practice, we would like to provide a deep insight about this kind of distinguishing attack when applied to the WG stream cipher and propose effective countermeasures in this contribution. To this end, we first extend Orumiehchiha *et al.*'s distinguishing attack to a much more general setting by building a k-order distinguisher. We then characterize some criteria for protecting the WG stream cipher family from the general distinguishing attack. Our analysis shows that by properly selecting the minimal polynomial of the LFSR for a WG stream cipher, the general distinguishing attack can be easily defeated.

The rest of the paper is organized as follows. In Section 2, we give a brief introduction about the WG stream cipher and the distinguishing attack proposed in [7]. The generalization of the Orumiehchiha *et al.*'s distinguishing attack will be addressed in Section 3. The criteria for protecting the WG stream ciphers from this attack is discussed in Section 4.

2 Preliminaries

In this section, we will introduce necessary definitions and results which will be used later. For the definitions and results of linear register feedback sequences, please refer to [1]. The following notations will be used throughout the paper.

- We denote a finite field $GF(2^n)$ as $\mathbb{F}_{2^n}$, and $\mathbb{F}_{2^n}^*$, the multiplicative group of $\mathbb{F}_{2^n}$.
- $\mathbb{F}_2^n = \{(x_0, \cdots, x_{n-1}) \,|\, x_i \in \mathbb{F}_2\}$.
- $\mathbf{a} = \{a_i\}$, a sequence over $\mathbb{F}_2$, is called a binary sequence. If $\mathbf{a}$ is a periodic sequence with period v, then we also denote $\mathbf{a} = (a_0, \cdots, a_{v-1})$ an element of $\mathbb{F}_2^v$.

2.1 WG Permutation and WG Transformation

Let $t(x) = x + x^{q_1} + x^{q_2} + x^{q_3} + x^{q_4}$ be a function defined on $\mathbb{F}_{2^m}$, where

$$\begin{aligned} q_1 &= 2^k + 1 \\ q_2 &= 2^{2k} + 2^k + 1 \\ q_3 &= 2^{2k} - 2^k + 1 \\ q_4 &= 2^{2k} + 2^k - 1 \end{aligned} \tag{1}$$

for k such that $3k \equiv 1 \mod m$. Then the function $\mathrm{WGP} : \mathbb{F}_{2^m} \to \mathbb{F}_{2^m}$ defined by

$$\mathrm{WGP}(x) = t(x+1) + 1, \ x \in \mathbb{F}_{2^m}$$

is called the *WG permutation*, and the Boolean function $\mathrm{WG} : \mathbb{F}_{2^m} \to \mathbb{F}_2$ defined by

$$\mathrm{WG}(x) = \mathrm{Tr}\left(t(x+1) + 1\right)$$

is called the *WG transformation* of $\mathrm{Tr}(t(x))$, or the *WG transformation* for short. Note that the WG permutations and transformations exist only if $m \,(\mathrm{mod}\, 3) \neq 0$. Moreover, let θ be a primitive element of $\mathbb{F}_{2^m}$, define two sequences $\mathbf{a} = \{a_i\}$ and $\mathbf{b} = \{b_i\}$ as

$$a_i = \mathrm{Tr}(t(\theta^i)) \text{ and } b_i = \mathrm{WG}(\theta^i) = \mathrm{Tr}\left(t(\theta^i + 1) + 1\right).$$

Then $\mathbf{b}$ is called a *WG transformation sequence of* $\mathbf{a}$, or a *WG sequence* for short. The WG sequences have many good properties, for instance ideal two-level autocorrelation, ideal 2-tuple distribution, balancedness, etc. Interested readers may refer to [2] for more details on WG transformations and WG sequences.

2.2 Description of WG Cipher

A WG cipher can be regarded as a nonlinear filter generator over an extension field [6]. A WG cipher, as shown in Figure 1, consists of a linear feedback shift register (LFSR), followed by a WG transform. The LFSR generates an m-sequence over $\mathbb{F}_{2^m}$ with period $2^n - 1$ where $n = ml$ and the connecting polynomial is a primitive polynomial $p(x)$ over $\mathbb{F}_{2^m}$ with degree l with $p(x) = x^l + \sum_{i=0}^{l-1} c_i x^i, c_i \in \mathbb{F}_2, c_0 \in \mathbb{F}_{2^m}$. The feedback signal Init is used only in the initialization phase of operation. When the cipher is running, the only feedback is within the LFSR and the output of the cipher is one bit per clock cycle. We denote a WG generator with an LFSR of l stages over $\mathbb{F}_{2^m}$ as an $WG(m, l)$ generator.

The mathematical expressions of updating the LFSR and the output sequence of the $WG(m, l)$ generator are given by

Update of LFSR:

$$\begin{aligned} a_{k+l} &= \begin{cases} \sum_{i=0}^{l-1} c_i a_{i+k} + \mathrm{WGP}(a_{k+l-1}) & 0 \leq k < 2l \text{ (in initialization phase)} \\ \sum_{i=0}^{l-1} c_i a_{i+k} & k \geq 2l \text{ (in running phase)} \end{cases} \\ \text{Output:} \quad s_k &= \mathrm{WG}(a_{k+2l+l-1}), k = 0, 1, \cdots \end{aligned} \tag{2}$$

where $\mathrm{WG}(\cdot)$ is defined in Section 2.1.

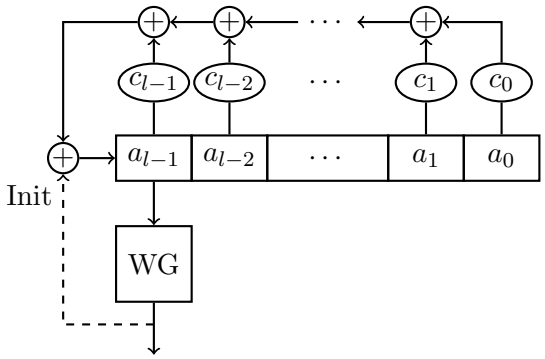

Figure 1: A general structure of $WG(m, l)$ generator

Particularly, for the WG-7 cipher WG(7, 23) in [5], the characteristic polynomial $p(x) = x^{23}+x^{11}+\beta \in \mathbb{F}_{2^7}[x]$, and the finite field $\mathbb{F}_{2^7}$ is generated by the primitive polynomial x^7+x+1 over $\mathbb{F}_2$.

2.3 A Distinguishing Attack on WG(7, 23)

Recently, Orumiehchiha, Pieprzyk, and Steinfeld [7] discovered a distinguishing attack on $WG(7, 23)$. The main idea is to use the linear relationship among the terms of the m-sequence $\{a_i\}$ generated by the LFSR to build the approximated linear relationship of the WG sequence $s_i = \mathrm{WG}(a_i), i \geq 0$ (here the initialization process is ignored). Recall that, for $WG(7, 23)$ in [5], the characteristic polynomial of the LFSR $p(x) = x^{23} + x^{11} + \beta \in \mathbb{F}_{2^7}[x]$. Two approximations of the linear equations used in [7] are listed below.

	Linear relation of$\{a_i\}$	Imbalanced of
(1)	$a_{23+i} + a_{11+i} + \beta a_i = 0,$	$s_{23+i} + s_{11+i} + s_i$ $= \mathrm{WG}(a_{11+i} + \beta a_i) + \mathrm{WG}(a_{11+i}) + \mathrm{WG}(a_i)$ 14 variables
(2)	$a_{23\cdot 2^7+i} + a_{11\cdot 2^7+i} + \beta a_i = 0$ (by $\beta^{2^7} = \beta$)	
(1) + (2)	$a_{23\cdot 2^7+i} + a_{11\cdot 2^7+i} + a_{23+i} + a_{11+i} = 0$	
	$a_{23\cdot 2^7-11+i} + a_{11\cdot 2^7+i-11} + a_{12+i} + a_i = 0$	$s_{23\cdot 2^7-11+i} + s_{11\cdot 2^7+i-11} + s_{12+i} + s_i$ $= \mathrm{WG}(a_{11\cdot 2^7-11+i} + a_{12+i} + a_i)$ $+\mathrm{WG}(a_{11\cdot 2^7-11+i}) + \mathrm{WG}(a_{12+i}) + \mathrm{WG}(a_i)$ 21 variables

We define

$$\begin{aligned} F_1(a_{11+i}, a_i) &= \mathrm{WG}(a_{11+i} + \beta a_i) + \mathrm{WG}(a_{11+i}) + \mathrm{WG}(a_i), \text{ and} \\ F_2(a_{1397+i}, a_{12+i}, a_i) &= \mathrm{WG}(a_{1397+i} + a_{12+i} + a_i) + \mathrm{WG}(a_{1397+i}) + \mathrm{WG}(a_{12+i}) + \mathrm{WG}(a_i), \end{aligned}$$

where $1397 = 11 \times 127$ and $i \geq 0$ is an integer. In [7], their respective probabilities of $F_1 = 0$ and $F_2 = 0$ are given by

$$\begin{aligned} \mathrm{P}\{F_1(a_{11+i}, a_i) = 0\} &= \tfrac{1}{2} + 2^{-7.415}, \\ \mathrm{P}\{F_2(a_{1397+i}, a_{12+i}, a_i) = 0\} &= \tfrac{1}{2} + 2^{-6.78}. \end{aligned}$$

Note. In [7], there is a typo in the bias of $\mathrm{P}\{F_1(a_{11+i}, a_i) = 0\}$, which should be $2^{-7.415}$ as shown above.

Remark 1 *It is mentioned in Section 2.1 that the sequence generated by* $\mathrm{WG}(x)$ *is balanced and has ideal* 2*-level autocorrelation, which implies that both* $\{s_i\}$ *and* $\{s_i + s_{i+\tau}\}$ *are balanced for any nonzero* τ*. But the sum of more than two* WG *sequences could be either balanced or imbalanced.*

2.4 Two Important Lemmas

In this subsection we list two important lemmas which will be used in the next two sections. The first result can be found in [9]. For the convenience of the reader, we provide a proof in the Appendix.

Lemma 1 *Let* **a** *be a sequence generated by a truly random generator with n bits, then the imbalance of* **a** *is approximately to $\sqrt{n/2\pi}$.*

The next result may be found in [8].

Lemma 2 *Given two binary random sequences, where the first is uniform and the other is biased, i.e. one binary value occurs with the probability $1/2+\epsilon$ while the other with the probability $1/2-\epsilon$. Then we need to observe $O\left(\frac{1}{\epsilon^2}\right)$ bits in order to distinguish the two distributions with a non-negligible probability of success.*

3 High Order Distinguisher from Linear Relations of the Recursive Relation of LFSRs

In [7], the linear approximations of WG-7 were obtained from the characteristic polynomial $p(x) = x^{23} + x^{11} + \beta, \beta \in \mathbb{F}_{2^7}$ of the LFSR. However, this attack can be generalized to the case that the characteristic polynomial has more than three terms. In this section, we will introduce a general distinguisher which exists for any filtering generator with or without memory.

3.1 Linear Relation of m-sequences

Let $p(x) = x^l + \sum_{i=0}^{l-1} c_i x^i \in \mathbb{F}_{2^m}[x]$ be a primitive polynomial of degree l. The following two properties for finite fields and m-sequences can be easily derived from the theory of finite fields [4] and m-sequences [1].

Property 1 *Let α be a root of $p(x)$ in its splitting field $\mathbb{F}_{2^{ml}}$. Then for any integer $0 \le k \le l-1$ and $0 \le i_0 < \cdots < i_{k-1} < 2^{ml}-1$ and $t_j \in \mathbb{F}_{2^m}$, if $g(\alpha) \neq 0$ where*

$$g(x) = \sum_{j=0}^{k-1} t_j x^{i_j}$$

then there exists an integer τ with $0 \le \tau < 2^{ml}-1$ such that $\alpha^\tau = g(\alpha)$.

Proof. The result follows from the fact that $g(\alpha)$ is an element of $\mathbb{F}_{2^{ml}}$ and α is a generator of the multiplicative group of $\mathbb{F}_{2^{ml}}$. □

Property 2 *Let $\{a_i\}$ be an m-sequence generated by $p(x)$. Then there exists some τ: $0 \le \tau \le 2^{ml}-2$ such that*

$$a_i = \mathrm{Tr}(\alpha^\tau \alpha^i),\ i = 0, 1, \cdots .$$

From Properties 1 and 2, the following results are followed immediately.

Property 3 *With the notation in Property 1, we have*

$$a_{j+\tau} = t_0 a_{j+i_0} + \cdots + t_{k-1} a_{j+i_{k-1}}, j = 0, 1, \cdots . \tag{3}$$

Or equivalently,

$$a_{j+\tau} = L(a_{j+i_0}, \cdots, a_{j+i_{k-1}}), j = 0, 1, \cdots \tag{4}$$

where

$$L(\mathbf{x}) = \sum_{j=0}^{k-1} t_j x_j, \ \mathbf{x} = (x_0, \cdots, x_{k-1}).$$

The relation given by (3) is referred to as a *k-order linear relation of a remote term* of the LFSR, which is determined by the minimal polynomial $p(x)$ of $\mathbf{a} = \{a_i\}$. The following result is from ideal k-tuple distribution of m-sequences (see [1]).

Property 4 *Let* $\mathbf{a} = \{a_i\}, a_i \in \mathbb{F}_{2^m}$ *be an m-sequence generated by* $p(x)$ *of degree* l *and* $0 \leq i_0 < \cdots < i_{k-1} < \frac{2^{ml}-1}{2^m-1}$ *such that* $\{a_{j+i_v}\}, v = 0, \cdots, k-1$*, regarded as a vector of* $\mathbb{F}_2^N$ *where* $N = 2^{ml} - 1$*, are linear independent over* $\mathbb{F}_2$*. Then for any* $(b_0, \cdots, b_{k-1}) \in \mathbb{F}_{2^m}^k, 1 \leq k \leq l$*, the following results hold:*

$$\begin{aligned} |\{j \,|\, 0 \leq j < 2^{ml} - 1, \ (a_{j+i_0}, \cdots, a_{j+i_{k-1}}) = (b_0, \cdots, b_{k-1})\}| &= (2^m)^{l-k}, \\ |\{j \,|\, 0 \leq j < 2^{ml} - 1, \ (a_{j+i_0}, \cdots, a_{j+i_{k-1}}) = (0, \cdots, 0)\}| &= (2^m)^{l-k} - 1. \end{aligned}$$

Note that the independent condition in Property 4 is equivalent to saying that $\sum_{v=0}^{k-1} \alpha^{i_v} \neq 0$.

3.2 k-Order Correlation of Boolean Functions

Let $f(x)$ be a function from $\mathbb{F}_{2^m}$ to $\mathbb{F}_2$ with $f(0) = 0$ and $s_i = f(a_{i+v}), i = 0, 1, \cdots$, where $0 \leq v < l$. When f is the WG transform, we have $v = l - 1$ as specified in Section 2.2. Now, for any $\mathbf{t} = (t_0, \cdots, t_{k-1}) \in (\mathbb{F}_{2^m}^*)^k$ and $\mathbf{u} = (u_0, \cdots, u_{k-1}) \in (\mathbb{F}_{2^m}^*)^k$, define a function $F_{(\mathbf{t},\mathbf{u})} : \mathbb{F}_{2^m}^k \to \mathbb{F}_2$ by

$$F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = f(t_0 x_0 + \cdots + t_{k-1} x_{k-1}) + f(u_0 x_0) + \cdots + f(u_{k-1} x_{k-1}). \tag{5}$$

A *k-order correlation of f at* $(\mathbf{t}, \mathbf{u}) = (t_0, \cdots, t_{k-1}, u_0, \cdots, u_{k-1})$ is defined as

$$\Delta_f(t_0, \cdots, t_{k-1}, u_0, \cdots, u_{k-1}) = \sum_{(x_0, \cdots, x_{k-1}) \in \mathbb{F}_{2^m}^k} (-1)^{F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1})}. \tag{6}$$

We define the following two subsets of $\mathbb{F}_{2^m}^k$

$$\begin{aligned} A(\mathbf{t}, \mathbf{u}) &= |\{(x_0, \cdots, x_{k-1}) \in \mathbb{F}_{2^m}^k \mid F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 0\}|, \\ D(\mathbf{t}, \mathbf{u}) &= |\{(x_0, \cdots, x_{k-1}) \in \mathbb{F}_{2^m}^k \mid F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 1\}|. \end{aligned}$$

Clearly we have $A(\mathbf{t}, \mathbf{u}) + D(\mathbf{t}, \mathbf{u}) = 2^{mk}$ and then

$$\Delta_f(\mathbf{t}, \mathbf{u}) = A(\mathbf{t}, \mathbf{u}) - D(\mathbf{t}, \mathbf{u}) = 2A(\mathbf{t}, \mathbf{u}) - 2^{mk}. \tag{7}$$

The probability that $F_{(\mathbf{t},\mathbf{u})}$ is equal to zero is

$$\mathrm{P}\{F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 0\} = \frac{A(\mathbf{t}, \mathbf{u})}{2^{mk}}. \tag{8}$$

Using (7), we have

$$\mathrm{P}\{F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 0\} = \frac{1}{2} + \frac{\Delta_f(\mathbf{t}, \mathbf{u})}{2^{mk+1}}. \tag{9}$$

Remark 2 We consider the Boolean function $F_{(\mathbf{t},\mathbf{u})}$ as a random Boolean function with mk variables, and we may regard it as a binary sequence of length 2^{mk}. By Lemma 1, the expectation of the random variable $F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 0$ is given by

$$\mathrm{E}\left(P\{F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = 0\}\right) = \frac{1}{2} \pm \frac{\sqrt{2\pi}}{2\sqrt{2^{mk}}}.$$

3.3 k-Order Distinguisher

From Property 3, we can build the following distinguisher for a pseudo-random sequence $\{s_i\}$.

***k*-order Distinguisher:** Assume that

$$a_{j+\tau} = \sum_{v=0}^{k-1} t_v a_{j+i_v} \text{ where } 0 \le i_0 < i_1 < \cdots < i_{k-1} < 2^{mk} - 1, k < l, t_j \in \mathbb{F}_{2^m}^*. \tag{10}$$

We define

$$F(a_{j+i_0}, a_{j+i_1}, \cdots, a_{j+i_{k-1}}) = f(a_{j+\tau}) + \sum_{v=0}^{k-1} f(a_{j+i_v}). \tag{11}$$

The following distinguisher

$$\mathrm{P}\{F(a_{j+i_0}, a_{j+i_1}, \cdots, a_{j+i_{k-1}}) = 0\} = \frac{1}{2} \pm \epsilon,\ 0 < \epsilon < 1/2 \tag{12}$$

is called a *k-order distinguisher*.

From the definition of $F_{(\mathbf{t},\mathbf{u})}$ in (5), we have

$$F(a_{j+i_0}, a_{j+i_1}, \cdots, a_{j+i_{k-1}}) = F_{(\mathbf{t},\mathbf{u})}(a_{j+i_0}, a_{j+i_1}, \cdots, a_{j+i_{k-1}})$$

where $\mathbf{t} = (t_0, \cdots, t_{k-1}) \in (\mathbb{F}_{2^m}^*)^k$ and $\mathbf{u} = (1, \cdots, 1) \in \mathbb{F}_2^k$. The bias ϵ can be determined through the computation of the k-order correlation of f at $(\mathbf{t}, \mathbf{u})$, i.e. $\Delta_f(\mathbf{t}, \mathbf{u})$ under some conditions. (Note. The following result is a general case of cross correlation of two geometry sequences in [3].)

Proposition 1 *If $(i_0, \cdots, i_{k-1})$ satisfies the condition in Property 4, then*

$$\epsilon \approx \frac{\Delta_f(\mathbf{t}, \mathbf{u})}{2^{mk+1}}$$

where $\mathbf{t} = (t_0, \cdots, t_{k-1})$ and $\mathbf{u} = (1, \cdots, 1)$.

Proof. Let

$$\Delta = \sum_{j=0}^{2^{ml}-2} (-1)^{f(t_0 a_{j+i_0} + \cdots + t_{k-1} a_{j+i_{k-1}}) + \sum_{v=0}^{k-1} f(a_{j+i_v})}. \tag{13}$$

Using a similar approach as for $\Delta_f(\mathbf{t})$, we get

$$\epsilon = \frac{\Delta}{2(2^{ml} - 1)}. \tag{14}$$

Denote $\mathbf{z} = (\mathbf{t}, \mathbf{u}) = (t_0, \cdots, t_{k-1}, 1, \cdots, 1)$ for simplicity. Since $(i_0, \cdots, i_{k-1})$ satisfies the condition in Property 4, by Property 4, then we may use the substitution $a_{j+i_t} \leftrightarrow x_t$ as shown below.

$$\begin{aligned} \Delta &= 2^{m(l-k)} \sum_{(x_0, \cdots, x_{k-1}) \in \mathbb{F}_{2^m}^k} (-1)^{F_{\mathbf{z}}(x_0, \cdots, x_{k-1})} - 1 \\ &= 2^{m(l-k)} \Delta_f(\mathbf{z}) - 1 \end{aligned}$$

where the constant -1 comes from $(x_0, \cdots, x_{k-1}) = (0, \cdots, 0)$ occurs $2^{m(l-k)} - 1$ times and $f(0) = 0$. Substituting the above identity into (14), we have

$$\begin{aligned} \epsilon &= \frac{2^{m(l-k)}\Delta_f(\mathbf{z})-1}{2(2^{ml}-1)} \\ &= \frac{\Delta_f(\mathbf{z})}{2^{mk+1}-1/2^{m(l-k)-1}} - \frac{1}{2^{ml+1}-2} \\ &\approx \frac{\Delta_f(\mathbf{z})}{2^{mk+1}}. \end{aligned}$$

Thus, the assertion is established. □

From Proposition 1, the bias of F is computed through

$$F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = f(\sum_{i=0}^{k-1} t_i x_i) + \sum_{i=0} f(x_i), x_i \in \mathbb{F}_{2^m} \tag{15}$$

provided the condition of Property 4.

According to Lemma 2, we have the following property.

Proposition 2 *Let $f(x) = WG(x)$. Then the k-order distinguisher of $WG(m, l)$ needs $O(\epsilon^{-2})$ bits to distinguish the key stream of $WG(m, l)$ cipher from a truly random sequence generator. Furthermore, $WG(m, l)$ generator needs to generate at least τ bits at one session of the encryption.*

Proof. Let $y_j = F(a_{j+i_0}, a_{j+i_1}, \cdots, a_{j+i_{k-1}})$. For $P\{y_j = 0\} = \frac{1}{2} \pm \epsilon$, it needs to observe $O(\epsilon^{-2})$ bits of $\{y_j\}$ to distinguish $\{s_j\}$ from a truly random sequence. However, in order to compute y_j it needs to know the remote bit $s_{j+\tau}$ and $(s_{j+i_0}, \cdots, s_{j+i_{k-1}})$ for computing the bias of the distinguisher. In order to get $s_{j+\tau}$, the generator has to generate $s_0, \cdots, s_{j+\tau-1}$ for one j. □

Note that $O(\epsilon^{-2})$ bits can be collected from different sessions, and they may not be consecutive. How to estimate $\Delta_f(1, \cdots, 1)$ is a hard problem. However, for small m and k, it can be determined through computation.

Remark 3 Theoretically, the k-order distinguisher can use k-order correlation of f at $(\mathbf{t}, \mathbf{u})$ for any $\mathbf{u} \in (\mathbb{F}_{2^m}^*)^k$ instead of $\mathbf{u} = (1, \cdots, 1)$ as defined in (11). However, this type of distinguisher cannot be constructed as explained below. From (3) in Property 3, since $t_j \in \mathbb{F}_{2^m}$, together with the trace representation of m-sequences in Property 2, we may rewrite as follows

$$u_v a_{j+i_v} = a_{j+\tau_v},\ 0 \le v < k \tag{16}$$

where $\tau_v = \frac{2^{ml}-1}{2^m-1}e_v + i_v$ where $0 \le e_v < 2^m - 1, v = 0, \cdots, k-1$. In this case, the bias of F cannot be computed by the method of Proposition 1. On the other hand, the k-order distinguisher $F_{(\mathbf{t},\mathbf{u})}$ requests to obtain the remote terms of the key stream bits $s_{j+\tau}, s_{j+\tau_0}, \cdots, s_{j+\tau_{k-1}}$, which demands that the generator should generate at least $\frac{2^{ml}-1}{2^m-1}$ bits in one session. For example, in $WG(7, 23)$, it requests that $WG(7, 23)$ generates at least $\frac{2^{161}-1}{2^7-1} \approx 2^{154}$ bits in one session, which is infeasible in practice.

Example 1 *The distinguishers F_1 and F_2 on $WG(7, 23)$.*

1. The distinguisher F_1 on WG(7, 23), defined in Section 2.3 is the 2-order distinguisher. In other words, the bias of F_1 is computed through

   $$F_{(\beta,1,1,1)}(x_0, x_1) = WG(x_1 + \beta x_0) + WG(x_0) + WG(x_1), x_i \in \mathbb{F}_{2^7}$$

   where $(t_0, t_1) = (\beta, 1)$ and $\mathbf{u} = (1, 1)$ in [7], which has the bias $2^{-7.415}$.

   We can easily verify that the bias of

   $$F_{(\beta,1,\beta,1)} = WG(x_1 + \beta x_1) + WG(x_1) + WG(\beta x_0), x_i \in \mathbb{F}_{2^7}$$

   is $2^{-6.299}$ which is larger than $F_{(\beta,1,1,1)}$. However, this function does not satisfy the condition of Property 4, thus Proposition 1 cannot be applied. In other words, the bias of

   $$F(a_j, a_{j+11}) = WG(a_{j+\tau}) + WG(\beta a_j) + WG(a_{j+11}), \forall j \ge 0$$

   where $a_{j+\tau} = \beta a_j + a_{j+11}$ cannot be computed through the bias of $F_{(\beta,1,\beta,1)}(x_0, x_1)$. On the other hand, in order to have the distinguisher, it also requests that $WG(7, 23)$ generates more 2^{154} bits in one session, as discussed in Remark 3, which is infeasible.

2. For the distinguisher F_2 on WG(7, 23), because $\{a_j\}_{j\ge0}$, $\{a_{j+12}\}_{j\ge0}$, and $\{a_{j+1397}\}_{j\ge0}$ are linear independent, through mapping

   $$\begin{aligned} a_j &\longleftrightarrow x_0 \\ a_{j+12} &\longleftrightarrow x_1 \\ a_{j+1397} &\longleftrightarrow x_2 \end{aligned}$$

   then $k = 3$ and the bias of the 3-order distinguisher can be computed in terms of the bias

   $$F_{(\mathbf{t},\mathbf{u})}(x_0, x_1, x_2) = WG(x_0 + x_1 + x_2) + WG(x_0) + WG(x_1) + WG(x_2), x_i \in \mathbb{F}_{2^7}$$

   where $\mathbf{t} = \mathbf{u} = (1, 1, 1)$, which has $3m = 21$ variables.

4 Resiliency Characteristic Polynomials Against k-Order Distinguisher Attacks

Theoretically we can always build a k-order distinguisher as stated in Section 3. However, in practice, this type of the distinguishers can be easily defeated by selecting $p(x)$ such that either τ is over the current computational technology or $\Delta_f(t_0, \cdots, t_{k-1}, 1, \cdots, 1)$ is large, or equivalently, the bias is small.

4.1 Resiliency Condition

According to Proposition 1, for WG(m, l), in order to satisfy the linear independent condition, there are two cases.

Case 1. If $0 \le i_0 < \cdots < i_{k-1} < l$, then the linear independent condition is true.

Case 2. If there exist some $i_v \ge l$, then we need to check whether $\sum_{v=0}^{k-1} \alpha^{i_v} = 0$.

We may select $p(x)$ to satisfy the following conditions where α is a root of $p(x)$ in $\mathbb{F}_{2^{ml}}$.

Resiliency Condition for $p(x)$: Choosing $p(x)$ such that there is no k-order linear relation of a remote term for $\tau \le \Delta_1$ and $k \le \Delta_2$, where Δ_1 is determined by the computational power of the current technology, say 2^{40} for a moderate computer, and Δ_2, by the bias ϵ or equivalently, $\Delta_f(\mathbf{t}, 1, \cdots, 1)$ in (11). In other words, we have the following two conditions.

Case 1. For $k = 2, \cdots, \Delta_2$, any $t_i \in \mathbb{F}_{2^m}^*$, and any $\{i_0, \cdots, i_{k-1}\} \subset \{0, 1, \cdots, l-1\}$, if

$$\alpha^\tau = t_0\alpha^{i_0} + \cdots + t_{k-1}\alpha^{i_{k-1}} \ne 0,$$

then

$$\tau > \Delta_1 = 2^{40}.$$

Case 2. For $k = 2, \cdots, \Delta_2$, any $t_i \in \mathbb{F}_{2^m}^*$, any $\{i_0, \cdots, i_{k-1}\} \subset \{0, 1, \cdots, \Delta_1\}$ with some $l \le i_v < \frac{2^{ml}-1}{2^m-1}$, if

$$\alpha^\tau = t_0\alpha^{i_0} + \cdots + t_{k-1}\alpha^{i_{k-1}} \ne 0 \text{ and } \alpha^{i_0} + \cdots + \alpha^{i_{k-1}} \ne 0,$$

then

$$\tau > \Delta_1 = 2^{40}.$$

A primitive polynomial $p(x)$ satisfies the above two conditions is called a *resilient feedback polynomial of* $WG(m,l)$ with respect to the threshold values (Δ_1, Δ_2).

Remark 4 Except for the above resiliency condition, k can be bounded by the following three cases. (1) Against exhaustive key search attack (EKS): Let K be the number of key bits. Then k can be bounded by $km < K$. Otherwise, in order to compute $\Delta_{WG}(1, \cdots, 1)$, it needs to compute the exponential sum of 2^{km} terms. If it is computable, then it could do exhaustive search for the key with 2^K cases. (2) Against time-memory-data trade-off attack (TMD): $km < K/2$. Otherwise, the attacker will launch TMD attack instead of k-order distinguish attack. (3) Against linear span attack (LSA): $2^{km} > LS$ where LS is the linear span of $WG(m,l)$ sequences.

Remark 5 The conditions on (τ, k) implies that the characteristic polynomial used in $\mathrm{WG}(m,l)$ should have at least $(k+1)$ terms. Otherwise, it has the k-order distinguisher whose bias is computed by the bias of

$$F_{(\mathbf{t},\mathbf{u})}(x_0, \cdots, x_{k-1}) = \mathrm{WG}\left(\sum_{i=1}^{k-1} x_i + \gamma x_0\right) + \mathrm{WG}(x_0) + \cdots + \mathrm{WG}(x_{k-1}),$$

where $\{0 = i_0, \cdots, i_{k-1}\} = \{0 < i < l \,|\, c_i = 1\}$, $p(x) = x^l + \sum_{i=1}^{l-1} c_i x^i + \gamma, c_i \in \mathbb{F}_2, \gamma \in \mathbb{F}_{2^m}$, $x_v \longleftrightarrow a_{j+i_v}$ and $\mathbf{t} = (t_0, \cdots, t_{k-1}) = (\gamma, 1, \cdots, 1)$ and $\mathbf{u} = (1, \cdots, 1)$.

4.2 WG-7 with Resiliency Condition

For $WG(7,23)$, we may choose $p(x) = x^{23} + \sum_{i=1}^{22} c_i x^i + \gamma, c_i \in \mathbb{F}_2, \gamma \in \mathbb{F}_{2^7}$, a primitive polynomial over $\mathbb{F}_{2^7}$ of degree 23, has 8 nonzero coefficients such that $\Delta_1 = 2^{40}$ and $\Delta_2 = 8$ (there are many such primitive polynomials found by testing). Then the best distinguisher that an attacker can build is a function with 56 variables, and key stream bits should be 2^{40} apart. However, in any practical applications that $WG(7,23)$ targeted for, it is infeasible to generate

$$s_{2^{40}+j}, s_{2^{40}+j-1}, \cdots, s_{22+j}, \cdots, s_j$$

for one j. Note that those bits cannot be collected from different communication sessions with different keys and IVs, because the attacker cannot distinguish the indexes of the key streams from different sessions. Thus, this type of distinguisher cannot be built. Even the generator can output more than 2^{40} bits, which make it possible to build this distinguisher. However, in

this case, $k > \Delta_2 = 8$. Thus, the time complexity for computing ϵ for distinguisher F in $m\Delta_2$ variables is more than 2^{63}, since $mk > m\Delta_2$, the time complexity for making the distinguisher work is more than 2^{63}, which is infeasible in terms of the current computing technology. Note that it can be checked for Case 1 in Proposition 2 up to $k = 6$. However, the complexity to check whether there is no k-order distinguisher to satisfy the resiliency condition in Case 2 is high.

5 Conclusion

In this article, we introduce a general k-order distinguisher to a filtering generator over extension fields, such as the WG stream cipher family. We provide an analytic method to compute bias of the distinguishers in terms of the filtering function. We provide the countermeasures to this type of distinguishers by properly choosing the minimal polynomial of the LFSR for a WG stream cipher. Thus, this type of distinguishing attacks can be easily defeated.

Acknowledgement

The authors wish to thank Dr. Yin Tan for his help for providing the proof for Lemma 1. The research is supported by NSERC SPG.

Appendix

Proof of Lemma 1. Let $\xi = |S_0 - S_1|$ be the imbalance of $\mathbf{a}$, where $S_0 = \sharp\{i \in [1..n] | a_i = 0\}$ and $S_1 = n - S_0$. Then ξ is a random variable with values in the range $0 \leq \xi \leq n$. In the following we only discuss the case n is an even integer. The arguments of the case n is odd is similar and we omit it here. Now assume $\xi = k$ and we get $k = |S_0 - S_1| = |2S_0 - n|$ and then $S_0 = \frac{n \pm k}{2}$. Clearly the equation is only meaningful when k is even. By the symmetry we only consider the case $S_0 = \frac{n+k}{2}$. Now we have

$$\Pr(\xi = k) = 2 \cdot \binom{n}{\frac{n+k}{2}} \cdot \frac{1}{2^n} = \frac{1}{2^{n-1}} \cdot \binom{n}{\frac{n+k}{2}}.$$

Therefore, the expectation of ξ is

$$E(\xi) = \sum_{k=0, k \ even}^{n} \Pr(\xi_k) k = \frac{1}{2^{n-1}} \sum_{k} \binom{n}{\frac{n+k}{2}} k.$$

By the fact that $\lim_{n\to\infty} \frac{1}{2^{n-1}} \sum_k \binom{n}{\frac{n+k}{2}} k = \sqrt{n/2\pi}$ (which can be shown by $\sum_k \binom{n}{\frac{n+k}{2}} k = \frac{n}{2}\binom{n}{n/2}$ and the sterling formula $n! \approx \sqrt{2\pi n}\left(\frac{n}{e}\right)^n$), we finish the proof. □

References

[1] S. Golomb and G. Gong, *Signal Design for Good Correlation*, Cambridge University Press, 2005.

[2] G. Gong and A. Youssef, "Cryptographic Properties of the Welch-Gong Transformation Sequence Generators", *IEEE Transactions on Information Theory*, Vol. 48, No. 11, pp. 2837-2846, November 2002.

[3] A. Klapper, A.H. Chan, and M. Goresky, "Cross-correlations of Linearly and Quadratically Related Geometric Sequences and GMW Sequences", *Discrete Applied Mathematics*, Vol. 46, No. 1, pp. 1-20, 1993.

[4] R. Lidl and H. Niederreiter, *Finite Fields*, Cambridge University Press, 1997.

[5] Y. Luo, Q. Chai, G. Gong and X. Lai, "A Lightweight Stream Cipher WG-7 for RFID Encryption and Authentication", *the Proceedings of Global Telecommunications Conference (GLOBECOM 2010)*, pp. 1-6, 2010.

[6] Y. Nawaz and G. Gong, "WG: A Family of Stream Ciphers with Designed Randomness Properties", *Information Sciences*, Vol. 178, No. 7, pp. 1903-1916, April 1, 2008.

[7] M. Orumiehchiha, J. Pieprzyk and R. Steinfeld, "Cryptanalysis of WG-7: A Lightweight Stream Cipher", *Cryptography and Communications*, Vol. 4, No. 3-4, pp. 277-285, 2012.

[8] I. Mantin and A. Shamir, "A Practical Attack on Broadcast RC4", *The 8th International Workshop on Fast Software Encryption - FSE'01*, LNCS 2355, M. Matsui (ed.), Berlin, Germany: Springer-Verlag, pp. 152-164. 2001.

[9] J. V. Uspensky, *Introduction to Mathematical Probability*, McGraw-Hill, New York, 1937.