Nonlinear Equivalence of Stream Ciphers

Sondre Rønjom[1] and Carlos Cid[2]

[1] Crypto Technology Group,
Norwegian National Security Authority,
Bærum, Norway
[2] Information Security Group,
Royal Holloway, University of London
Egham, United Kingdom

Abstract. In this paper we investigate nonlinear equivalence of stream ciphers over a finite field, exemplified by the pure LFSR-based filter generator over $\mathbb{F}_2$. We define a nonlinear equivalence class consisting of filter generators of length n that generate a binary keystream of period dividing $2^n - 1$, and investigate certain cryptographic properties of the ciphers in this class. We show that a number of important cryptographic properties, such as algebraic immunity and nonlinearity, are not invariant among elements of the same equivalence class. It follows that analysis of cipher-components in isolation presents some limitations, as it most often involves investigating cryptographic properties that vary among equivalent ciphers. Thus in order to assess the resistance of a cipher against a certain type of attack, one should in theory determine the weakest equivalent cipher and not only a particular instance. This is however likely to be a very difficult task, when we consider the size of the equivalence class for ciphers used in practice; therefore assessing the exact cryptographic properties of a cipher appears to be notoriously difficult.

Keywords: Stream ciphers, sequences, nonlinear equivalence.

1 Introduction

A stream cipher [8] is a type of encryption algorithm which encrypts individual alphabet elements of a plaintext, one at a time, with a time-varying transformation. Stream ciphers are very popular due to their many attractive features: they are generally fast, can usually be implemented efficiently in hardware, have no (or limited) error propagation, and are particularly suitable for environments where no buffering is available and alphabet-elements need to be processed individually.

It is very common to construct stream ciphers based on *linear feedback shift registers* (LFSRs). Besides their attractive implementation features, the rich algebraic structure often enables a more formal and detailed security analysis. A filter generator over $\mathbb{F}_2$ is perhaps a stream cipher in its simplest form, with a well-defined mathematical description: it consists of a sequence generator and a Boolean function which produce a keystream based on the state of the register.

S. Hong and T. Iwata (Eds.): FSE 2010, LNCS 6147, pp. 40–54, 2010.

The security of such a construction is highly reliant on both the properties of the sequence-generator, as well as the properties of the Boolean function.

Boolean functions play a very important role in stream cipher design and analysis (as well as in several other cryptographic primitives), and a significant amount of literature has been devoted to the study of cryptographic properties of Boolean functions. Cryptanalytic techniques that may exploit these properties include correlation attacks, algebraic attacks, inversion attacks, among others.

We note however that for several methods of analysis one often investigates the Boolean function in isolation from the associated sequence generator. For instance, the algebraic normal form of a Boolean function can be constructed and related properties such as algebraic immunity, algebraic degree, nonlinearity and correlation immunity, can be computed to derive the cipher's security. On the other hand, other types of attacks take advantage of certain properties of the sequence generator. For instance, the Hamming weight of a feedback polynomial should not be low in order to resist correlation attacks; likewise, to resist inversion attacks, the positions of the cipher's LFSR which a Boolean function taps from, should satisfy additional requirements.

In this paper, we discuss and attempt to combine the analysis of both the generator and the corresponding Boolean function. Such an approach has for instance been taken by the authors of [10], enabling a very efficient attack on a class of stream ciphers by identifying certain characteristic structures which are not evident from isolated analysis of the cipher components. Our main focus point is to investigate (nonlinear) equivalence of LFSR-based stream ciphers using basic properties of Galois fields and certain isomorphisms between the corresponding multiplicative groups. This can be seen as a way of constructing *isomorphic* ciphers (examples of cipher representations and isomorphisms were provided in [1,9]; the subject was discussed in detail in [2]).

We show here that important cryptographic properties such as nonlinearity and algebraic immunity are variant with respect to such an equivalence. The focal point of this paper is therefore: since there are many ciphers generating the same keystream, any cryptographic property should be defined with respect to the weakest equivalent cipher. However, without some type of provable construction, it seems difficult to assess the exact security of a filter generator for practical sizes, since the class of equivalent ciphers is very large in practice. For instance, there are about 2^{121} nonlinearly equivalent filter generators with an LFSR of length 128 over $\mathbb{F}_2$ generating a keystream of period $2^{128} - 1$. We note however that we are not concerned here with affine equivalences, as such equivalences are not particularly revealing in general.

This paper is organized as follows. In section 2 we present some basic definitions and define the notation used in the paper. In section 3, the basic principle of equivalence and change of basis is introduced, and in section 4 we introduce an equivalence class of filter generators with respect to a periodic sequence. In section 5 we explain how to determine equivalences realised as nonlinear polynomial functions, and in section 6 we reflect on some consequences for the design and cryptanalysis of LFSR-based stream ciphers.

2 Preliminaries

In this section we provide some definitions which are essential in our analysis; see [6] and [5] for a more detailed discussion of sequences over finite fields.

Let p be a prime, $q = p^n$, and let $\mathbb{F}_q$ denote the finite field with q elements. The *order* of an element $\alpha \in \mathbb{F}_q$ is the smallest positive integer k such that $\alpha^k = 1$, denoted by $\mathrm{ord}(\alpha)$. An element α with order $q-1$ is called a *primitive element* and its minimal polynomial $g_\alpha(x) \in \mathbb{F}_p[x]$ is called a *primitive polynomial*. The primitive elements are exactly the generators of $\mathbb{F}_q^*$, the multiplicative group consisting of the non-zero elements of $\mathbb{F}_q$.

If α is a primitive element of $\mathbb{F}_q$ and $\gcd(k, q-1) = 1$, then any element α^k is also primitive. In particular, the conjugates α^{p^i} of α are all primitive and form the roots of the primitive polynomial $g_\alpha(x) = \sum_{i=0}^{n-1}(x - \alpha^{p^i})$ of degree n over $\mathbb{F}_p[x]$. It follows that there are $\phi(q-1)$ primitive elements of $\mathbb{F}_q$, where ϕ denotes Euler's totient function, and that the number of primitive polynomials over $\mathbb{F}_p$ of degree n is given by $\phi(q-1)/n$.

If k divides n, then $p^k - 1$ divides $q - 1 = p^n - 1$, and it follows that there is an element $\beta \in \mathbb{F}_q$ with order $p^k - 1$. Furthermore, β is a primitive element of $\mathbb{F}_p(\beta) \simeq \mathbb{F}_{p^k} \subseteq \mathbb{F}_p(\alpha) \simeq \mathbb{F}_q$.

The *absolute trace* of an element $\beta \in \mathbb{F}_{p^k} \subseteq \mathbb{F}_q$ is given by

$$\mathrm{Tr}_1^k(\beta) = \sum_{i=0}^{k-1} \beta^{p^i},$$

where $\mathrm{Tr}_1^k(x)$ denotes the trace function from $\mathbb{F}_{p^k}$ to $\mathbb{F}_p$. We write $\mathrm{Tr}(x) = \mathrm{Tr}_1^n(x)$ when there is no room for confusion. If $\alpha \in \mathbb{F}_q$ is a primitive element, then $\{1, \alpha, \ldots, \alpha^{n-1}\}$ is a basis of $\mathbb{F}_q$ (when considered as a vector space over $\mathbb{F}_p$).

Let $\mathbf{s}$ denote a periodic sequence over $\mathbb{F}_p$ with period e dividing $q-1$, viewed as a vector of length $q-1$, and let $m(x) = \sum_{i=0}^{k} c_i x^i \in \mathbb{F}_p[x]$ be a monic polynomial of degree k. We say that the sequence $\mathbf{s}$ satisfies the linear recurrence defined by $m(x)$ if

$$c_0 a_t + c_1 a_{t+1} + \ldots + c_{k-1} a_{t+k-1} + a_{t+k} = 0,$$

for all $t \geq 0$. The *minimal polynomial* of $\mathbf{s}$ is the polynomial of least degree whose linear recurrence is satisfied by $\mathbf{s}$.

We say that a sequence $\mathbf{s}$ is *irreducible* if its minimal polynomial is irreducible over $\mathbb{F}_p$. A sequence $\mathbf{s}$ is *generated* by a polynomial $g(x) \in \mathbb{F}_p[x]$, if the minimal polynomial $m_{\mathbf{s}}(x)$ of $\mathbf{s}$ divides $g(x)$. Denote by $\Omega(g(x))$ the vector space spanned by the sequences generated by $g(x)$. If $g(x)$ is primitive, then $\Omega(g)$ contains $q-1$ cyclically equivalent sequences (in addition to the zero-sequence), and every non-zero sequence in $\Omega(g)$ has maximal period $q-1$. Such sequences are called *maximal sequences* (or *m-sequences*).

Let $\mathbf{s}$ be an m-sequence over $\mathbb{F}_p$ with minimal polynomial $m_{\mathbf{s}}(x)$ of degree n, and $\alpha \in \mathbb{F}_q$ be a root of $m_{\mathbf{s}}(x)$ (and thus $m_{\mathbf{s}}(x) = g_\alpha(x)$). Then $\mathbf{s}$ may be written over $\mathbb{F}_q$ in terms of the roots of $m_{\mathbf{s}}(x)$ as

$$s_t = \mathrm{Tr}(X\alpha^t) = \sum_{i=0}^{n-1} (X\alpha^t)^{p^i}, \quad t = 0, 1, 2, \ldots,$$

where $X \in \mathbb{F}_q^*$. Furthermore, the $q-1$ nonzero choices of $X \in \mathbb{F}_q^*$ result in the $q-1$ distinct shifts of the same m-sequence $\mathbf{s}$.

In the remaining of this paper, we will consider sequences defined over the field $\mathbb{F}_2$, that is, $p = 2$ and $q = 2^n$. It should be noted however that the analysis provided here can be extended trivially to sequences and filter generators over any prime extension $\mathbb{F}_{p^n}$.

Let $R = \mathbb{F}_2[x_0, x_1, \ldots, x_{n-1}]$ and J be the ideal of R generated by the set $\{x_i^2 + x_i\}_{(0 \le i < n)}$. Any Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be realised as a polynomial function $f(x_0, ..., x_{n-1}) \in R/J$. The algebraic degree of the Boolean function f is the highest degree of a monomial in f.

A pure filter generator over $\mathbb{F}_2$ consists of a LFSR-based sequence generator and a nonlinear Boolean function. Moreover, if we let $\mathbf{s}$ denote an m-sequence over $\mathbb{F}_2$ and $f \in R/J$ a nonlinear function, then a nonlinearly filtered sequence $\mathbf{a}$ may be generated by

$$a_t = f(s_t, s_{t+1}, \ldots, s_{t+n-1}), \ t = 0, 1, 2, \ldots,$$

for some initial state $(s_0, s_1, \ldots, s_{n-1})$ of the LFSR generating $\mathbf{s}$. It is well known that a filter generator with an LFSR of period e, can generate any sequence of period l dividing e for appropriate choices of filter function f.

3 Equivalent Sequence Generators

Our main motivation results from the following observation: an m-sequence $\mathbf{s}$ of period $q - 1 = 2^n - 1$ may in general be written in terms of the roots of *any* primitive polynomial of degree n in $\mathbb{F}_2[x]$.

Indeed, let $\beta = \alpha^k$ be a primitive element of $\mathbb{F}_2(\alpha) \simeq \mathbb{F}_q$. Then $\gcd(k, q-1) = 1$, and the k-power exponentiation is an automorphism of the multiplicative group $\mathbb{F}_q^*$. Furthermore, this automorphism induces the mapping $x^k : \mathbb{F}_2(\alpha) \to \mathbb{F}_2(\beta)$, with inverse x^r, where r is the multiplicative inverse of k modulo $q-1$.

Let $\mathbf{s} \in \Omega(g_\alpha(x))$ be an m-sequence generated by

$$s_t = \mathrm{Tr}(X\alpha^t), \tag{1}$$

and $\beta = \alpha^k \in \mathbb{F}_2(\alpha)$, where $\mathrm{ord}(\beta) = q-1$, and let $r \cdot k \equiv 1 \pmod{q-1}$. Then we may rewrite (1) in terms of the primitive element β as

$$s_t = \mathrm{Tr}(X\alpha^t) = \mathrm{Tr}((Y\beta^t)^r), \tag{2}$$

where $Y \in \mathbb{F}_2(\beta)$ and $Y^r = X$ is an elementary change of basis. Equation (2) shows how an m-sequence $\mathbf{s} \in \Omega(g_\alpha)$ may be represented (nonlinearly) in terms of the roots of the minimal polynomial of another m-sequence $\mathbf{b} \in \Omega(g_\beta)$. In particular, the output of the LFSR satisfying the linear recursion defined by $g_\alpha(x)$ may also be generated by a nonlinear filter generator using an LFSR satisfying the linear recursion defined by $g_\beta(x)$, as illustrated in the following example.

Example 1. Let $n = 5, q = 2^n = 32$ and let $\mathbb{F}_2(\alpha) \simeq \mathbb{F}_{32}$, where $g_\alpha(x) = x^5 + x^4 + x^3 + x^2 + 1 \in \mathbb{F}_2[x]$ is a primitive polynomial. An m-sequence $\mathbf{s} \in \Omega(g_\alpha(x))$ can be generated by

$$s_t = \mathrm{Tr}(X\alpha^t), t = 0, 1, 2, \ldots,$$

where $X \in \mathbb{F}_{32}^*$. Now let $\beta = \alpha^{21}$ and $X^{21} = Y \in \mathbb{F}_2(\beta)$. It follows that

$$\mathrm{Tr}(X\alpha^t) = \mathrm{Tr}((Y\beta^t)^3), t = 0, 1, 2, \ldots,$$

since $3 \cdot 21 \equiv 1 \pmod{31}$.

The corresponding filter generator over $\mathbb{F}_2(\beta)$ is given by

$$s_t = f(b_t, b_{t+1}, \ldots, b_{t+4}), t = 0, 1, 2, \ldots,$$

where

$$(b_t, b_{t+1}, \ldots, b_{t+4}) = (\mathrm{Tr}(Y\beta^t), \mathrm{Tr}(Y\beta^{t+1}), \ldots, \mathrm{Tr}(Y\beta^{t+4})),$$

and

$$f(x_0, x_1, x_2, x_3, x_4) = x_0x_2 + x_2x_3 + x_1x_4 + x_2x_4 + x_1 + x_3.$$

The two filter generators (one of them is linear) will generate identical sequences for all possible initial states X and $Y = X^{21}$, and they are thus equivalent sequence generators.

Notice that the function f has algebraic immunity 2 and nonlinearity 12, which is maximal for a quadratic Boolean function in 5 variables. Thus, on the basis of certain types of analysis, one of the ciphers appears to be secure while the other is not.

Example 1 illustrates that the Boolean function corresponding to the trace-representation over $\mathbb{F}_2(\beta)$ may possess strong cryptographic properties in general. Thus, if we investigate the security of the whole cipher by analysing the Boolean function of a particular filter generator in isolation, we might perhaps conclude (erroneously) that it is a cryptographically strong cipher.

4 Equivalence of Filter Generators

In order to simplify the presentation, we introduce the following notation.

Definition 1. *Let $X, \alpha \in \mathbb{F}_q^*$. Then define the vector*

$$S(X\alpha^t) = (Tr_1^k(X\alpha^t), Tr_1^k(X\alpha^{t+1}), \ldots, Tr_1^k(X\alpha^{t+k-1})) \in \mathbb{F}_2^k,$$

where $k = \dim(\mathbb{F}_2(\alpha))$ and k divides n.

The vector $S(X\alpha^t)$ is equivalent to the state at time t of an LFSR with characteristic polynomial $g_\alpha(x)$ of degree k and initial state $S(X) \in \mathbb{F}_2^k$. In the remaining of this paper, we will only consider the case $k = n$. We view any Boolean function in $r \leq n$ variables as a polynomial in $\mathbb{B}_n = R/J$. For convenience in the presentation, we have the following definition.

Definition 2. *Let $\beta, X \in \mathbb{F}_q^*$, $b_t = Tr(X\beta^t)$ be a linear recurrence sequence and $f \in \mathbb{B}_n$ a Boolean function. Define a sequence*

$$\mathcal{L}_\beta(f, t, X) = (f(S(X\beta^t)), f(S(X\beta^{t+1})), \ldots, f(S(X\beta^{t+q-2}))),$$

of length $q-1$, with entries

$$f(S(X\beta^t)) = f(b_t, b_{t+1}, \ldots, b_{t+n-1}).$$

Let $\mathcal{L}_\beta(f)$ be the set of sequences defined as

$$\mathcal{L}_\beta(f) = \{\mathcal{L}_\beta(f, 0, X) \mid X \in \mathbb{F}_q^*\}.$$

The set $\mathcal{L}_\beta(f)$ can be seen as the set of all possible keystream output sequences (of length $q-1$) from a filter generator, whose LFSR has feedback polynomial $g_\beta(x)$ and filtering function f. The non-zero elements $X \in \mathbb{F}_q^*$ determine the initial state of the LFSR.

The period of the sequences in $\mathcal{L}_\beta(f)$ depend on the order of β and the function f; for instance, it should be clear that the period of any sequence in $\mathcal{L}_\beta(f)$ cannot be greater than $\text{ord}(\beta)$, and in fact must divide $\text{ord}(\beta)$ (in particular, it divides $q-1$). In general, if $\text{ord}(\beta) = q-1$, then for a random function $f \in \mathbb{B}_n$, the sequence $\mathcal{L}_\beta(f)$ have almost surely period $q-1$.

When considering the set of (polynomial) Boolean functions $\mathbb{B}_n = R/J$, we can define a surjective homomorphism φ from $\mathbb{B}_n$ to the set of sequences over $\mathbb{F}_2$ of length $q-1 = 2^n - 1$ as

$$\begin{aligned} \varphi : \mathbb{B}_n &\to \mathbb{F}_2^{q-1} \\ f &\mapsto \mathbf{s}_f, \end{aligned}$$

where $\mathbf{s}_f$ corresponds to the truth-table of f in all points of $\mathbb{F}_2^n$ except $(0, \ldots, 0)$.

It can be shown that $ker(\varphi) = \langle h \rangle$, where $h(x_0, \ldots, x_{n-1}) = \prod_{i=0}^{n-1}(x_i + 1)$, and as a result $\varphi(f_1) = \varphi(f_2)$ if, and only if, $f_1 \equiv f_2 \bmod \langle h \rangle$. Moreover, since $\langle h \rangle = \{0, h\}$, we have the counter-image of any sequence $\mathbf{s} \in \mathbb{F}_2^{q-1}$ given by

$$\varphi^{-1}(\mathbf{s}) = \{f_\mathbf{s}, f_\mathbf{s}^*\} \subset \mathbb{B}_n, \tag{3}$$

where $f_\mathbf{s}^* = f_\mathbf{s} + h$. Note that $f_\mathbf{s}, f_\mathbf{s}^*$ are the functions that coincide in the set $\mathbb{F}_2^n \setminus \{(0, \ldots, 0)\}$ (with image in this set given by the sequence $\mathbf{s}$), but with $f_\mathbf{s}(0, \ldots, 0) \neq f_\mathbf{s}^*(0, \ldots, 0)$.

Definition 3. *For a sequence $\mathbf{s} \in \mathbb{F}_2^{q-1}$ and $\beta \in \mathbb{F}_q^*$, let*

$$V_\beta(\mathbf{s}) = \{f \in \mathbb{B}_n \mid \mathbf{s} \in \mathcal{L}_\beta(f)\}.$$

In other words, we can consider $V_\beta(\mathbf{s})$ as the set of all filter generators with feedback polynomial $g_\beta(x)$ that generate $\mathbf{s}$ as its first $q-1$ terms. The following lemma summarises the conjugation property of such sets.

Lemma 1. *For any $\mathbf{s} \in \mathbb{F}_2^{q-1}$ and $\beta \in \mathbb{F}_q^*$, we have*

$$V_{\beta^{2^i}}(\mathbf{s}) = V_{\beta^{2^j}}(\mathbf{s}), 0 \leq i, j \leq n-1.$$

The above lemma follows directly from the fact that $g_{\beta^{2^i}}(x) = g_{\beta^{2^j}}(x)$. We then have the following lemma.

Lemma 2. *Let* $\mathbf{s} \in \mathbb{F}_2^{q-1}$ *denote a periodic sequence with* $e = per(\mathbf{s})$ *and* $\beta \in \mathbb{F}_q^*$, *where* $per(\mathbf{s}) \mid ord(\beta)$. *Then*

$$|V_\beta(\mathbf{s})| \leq \frac{e(q-1)}{ord(\beta)} \cdot 2^{q-ord(\beta)}.$$

Proof. Let $w = \mathrm{ord}(\beta)$ and $\mathcal{X}$ be the subgroup of $\mathbb{F}_q^*$ generated by β. The subgroup $\mathcal{X}$ has index $k = (q-1)/w$ in $\mathbb{F}_q^*$, and thus there are k elements $1 = X_0, X_1, X_2, \ldots, X_{k-1} \in \mathbb{F}_q^*$ such that the sets $\mathcal{X}_i = X_i\mathcal{X}$ form a partition of $\mathbb{F}_q^*$ (these are the cosets of $\mathcal{X}$ in $\mathbb{F}_q^*$).

We can thus associate the sets $\mathcal{X}_i \subseteq \mathbb{F}_q^*$ with the distinct and non-intersecting *ordered* sets

$$V_i = \{S(X_i\beta^t) = v_i^t \mid t = 0, 1, \ldots, w-1\} \subseteq \mathbb{F}_2^n.$$

It is clear that the elements $X_0, X_1, \ldots, X_{k-1} \in \mathbb{F}_q^*$ result in the k distinct and shift-nonequivalent state-cycles of the corresponding LFSR with period w.

Let $H = \{h_0, h_1, \ldots, h_{k-1}\} \subseteq \mathbb{B}_n$ be the set of Boolean polynomials such that $h_i(x) = 0$ if $x \in V_i$ and $h_i(x) = 1$ if $x \in \mathbb{F}_2^n \backslash V_i$. The ideal $\langle h_i \rangle$ consists of the set of all functions in $\mathbb{B}_n$ that are zero when restricted to V_i. Since V_i has cardinality w, then $|\langle h_i \rangle| = 2^{q-w}$ for every i.

Given $\mathbf{s} \in \mathbb{F}_2^{q-1}$ with period e, for every V_i we can define the function $f_i \in \mathbb{B}_n$ as $f_i(v_i^t) = s_t$ for $0 \leq t \leq w-1$, and $f_i(x) = 0$ if $x \in \mathbb{F}_2^n \backslash V_i$. Thus $f_i \in V_\beta(\mathbf{s})$. Furthermore, it is clear that if $g_i \equiv f_i \mod \langle h_i \rangle$, then $g_i(v_i^t) = s_t$ for $0 \leq t \leq w-1$, and $g_i \in V_\beta(\mathbf{s})$.

Now, by considering the w shift-equivalent sets V_i' of the ordered set V_i, we obtain shift-equivalent functions to the elements of the set $F_i = \{f_i + g_i \mid g_i \in \langle h_i \rangle\} \subset \mathbb{B}_n$. In fact we get $e = per(\mathbf{s})$ such functions for each element in F_i. Thus, for every V_i, we have $e \cdot 2^{q-w}$ functions in $V_\beta(\mathbf{s})$. We can repeat the above with all k sets V_i to obtain

$$k \cdot e \cdot 2^{q-w} = \frac{e(q-1)}{w} \cdot 2^{q-w}$$

elements in $V_\beta(\mathbf{s})$. □

The inequality in lemma 2 is necessary in case $per(\mathbf{s}) < 2^n - 1$, since it may then be the case that some of the functions are counted several times. However, the main motivation of this paper is sequences of maximal period and one should note that when $per(\mathbf{s}) = \mathrm{ord}(\beta) = 2^n - 1$, then $|V_\beta(\mathbf{s})| = 2(q-1)$; in fact, we have that $\varphi^{-1}(\mathbf{s})$ contains the two representatives of the shift equivalence classes in $V_\beta(\mathbf{s})$ (assuming the natural ordering on the elements of $\mathbb{F}_2^n$ induced by the cyclic group generated by β). This fact also implies the following lemma.

Lemma 3. *Let* β *be a primitive element of* $\mathbb{F}_q$, *and* $f \in \mathbb{B}_n$. *If* $\mathbf{s}_1$ *and* $\mathbf{s}_2$ *are sequences in the set* $\mathcal{L}_\beta(f)$, *then* $V_\beta(\mathbf{s}_1) = V_\beta(\mathbf{s}_2)$.

We note that when β is not primitive, then lemma 3 is not necessarily true.

In the following definition, we assume sequences with period $e|(q-1)$, where e is not a divisor of $2^k - 1$, $0 < k < n$. That is, we assume that the sequences are generated by filter generators consisting of irreducible LFSRs of length n.

Definition 4. *Let $\mathbf{s} \in \mathbb{F}_2^{q-1}$ be a sequence with period e dividing $q-1$, where e is not a divisor of $2^k - 1$, with $0 < k < n$. Then let*

$$\mathbb{G}_n(\mathbf{s}) = \{V_\beta(\mathbf{s}) \,|\, \beta \in \mathbb{F}_q, \;\; e \,|\, ord(\beta)\}.$$

In other words, the set $\mathbb{G}_n(\mathbf{s})$ may be viewed as a class of filter generators of length n that generate $\mathbf{s}$ as a keystream. For sequences with period e dividing $q-1$, the size of $\mathbb{G}_n$ is given by the following proposition.

Proposition 1. *If $\mathbf{s} \in \mathbb{F}_2^{q-1}$ has period e dividing $q-1$, where e is not a divisor of $2^k - 1$, with $0 < k < n$, then*

$$|\mathbb{G}_n(\mathbf{s})| = \sum_{e|w} \phi(w)/n,$$

where the sum is extended over all positive divisors w of $q-1$.

Proof. By restricting the class $\mathbb{G}_n(\mathbf{s})$ to sequences with period e, where e is not a divisor of $2^k - 1$, with $0 < k < n$, we are restricting the sets V_β to elements β with minimal polynomial of degree n over $\mathbb{F}_2$. Thus, we need only count the distinct irreducible polynomials in $\mathbb{F}_2[x]$ of degree n with periods of which e is a divisor. □

The following corollary then follows immediately, which is of most interest for this paper.

Corollary 1. *If $\mathbf{s} \in \mathbb{F}_2^{q-1}$ has period $q-1$, then*

$$|\mathbb{G}_n(\mathbf{s})| = \phi(q-1)/n,$$

where $\phi(q-1)$ is the number of generators of the multiplicative group of $\mathbb{F}_q$.

Thus when $per(\mathbf{s}) = q-1$, the set $\mathbb{G}_n(\mathbf{s})$ contains $\phi(q-1)/n$ elements, where each element $V_\beta(\mathbf{s})$ contains two equivalent functions with respect to $\mathbb{F}_2^n \setminus \{0\}$ (without counting affine equivalences). There are thus in total

$$2 \cdot \phi(q-1)/n \tag{4}$$

distinct filter generators with feedback-polynomial of degree n that generate $\mathbf{s}$ (again, without counting affine equivalences).

Remark 1. Assume that we determine $\mathbb{G}_n$ for a sequence $\mathbf{s}$ of period $e < q-1$ and assume that the sequence stems from a filter generator with irreducible (but not primitive) feedback polynomial of degree n. Such filter generators (most often) produce $r = (q-1)/e$ *shift-nonequivalent* sequences of period e. Thus, the equivalence only encapsulates one out of $r = (q-1)/e$ sequences generated by that generator, and we are only guaranteed that the two generators are equivalent for a subset of initial states. Thus, $\mathbb{G}_n$ induce a *strong equivalence* for sequences with periods $2^k - 1$ (see Proposition 2), and a weak form of equivalence otherwise. This will be studied in closer detail in a follow-up paper.

We have restricted $\mathbb{G}_n(\mathbf{s})$ to the set of filter generators with feedback polynomial of degree n for the purpose of simplicity and clarity of the presentation. Our main focus are sequences with period $q-1$, the case of filter generators with a primitive feedback polynomial, in which the equivalence class $\mathbb{G}_n$ becomes especially simple and clear. While it is possible to generalise $\mathbb{G}_n$ into more complex equivalence classes offering more insight in cryptanalysis, it is out of the scope of this paper. In particular, one may generalise $\mathbb{G}_n$ by incorporating combiner generators that generate the same sequences or for instance filter generators based on NLFSRs. For instance, it should be clear that a sequence generated by a combiner generator, can also be generated by a filter generator, and vice-versa.

It is especially simple to deduce equivalent ciphers generating a sequence of period $q-1$ in terms of nonlinear equivalences of Boolean functions. In the following section, we describe how to deduce isomorphic filter generators in the case of sequences of period $q-1$.

5 Structure of Equivalent Functions

With access to a filter generator that generates a sequence $\mathbf{a}$, we may in fact generate all other equivalent filter generators.

Let $\mathbb{F}_2(\alpha) \simeq \mathbb{F}_q$ and let $\beta = \alpha^k$ be a primitive element of $\mathbb{F}_2(\alpha)$. Then for any elements $X \in \mathbb{F}_2(\alpha)$ and $Y \in \mathbb{F}_2(\beta)$, let $\phi_\beta(x_0, x_1, \ldots, x_{n-1})$ be the vectorial Boolean function which maps states $S(X\alpha^t) \in \mathbb{F}_2^n$ to states $S(Y\beta^t) \in \mathbb{F}_2^n$. Moreover, we have that

$$\phi_\beta(x_0, x_1, \ldots, x_{n-1}) = (y_0, y_1, \ldots, y_{n-1}),$$

and thus $\phi_\beta(S(X\alpha^t)) = S(Y\beta^t)$.

Now if we select a function $f_\alpha(x_0, x_1, \ldots, x_{n-1}) \in \mathbb{B}_n$, then we may compute another function by

$$\begin{aligned} f_\alpha(x_0, x_1, \ldots, x_{n-1}) &= f_\alpha(x_0, x_1, \ldots, x_{n-1}) \circ \phi_\beta^{-1}(y_0, y_1, \ldots, y_{n-1}) \\ &= f_\alpha(\phi_{\beta,0}^{-1}(y_0, \ldots, y_{n-1}), \ldots, \phi_{\beta,n-1}^{-1}(y_0, \ldots, y_{n-1})) \\ &= f_\beta(y_0, y_1, \ldots, y_{n-1}), \end{aligned}$$

where $\phi_\beta^{-1}(y_0, y_1, \ldots, y_{n-1})$ is the inverse of $\phi_\beta(x_0, x_1, \ldots, x_{n-1})$. And since $Y = X^k$, it follows that

$$a_t = f_\beta(S(Y\beta^t)) = f_\alpha(S(X\alpha^t)), t = 0, 1, 2, \ldots,$$

which corresponds to two filter generators with distinct LFSRs and filter functions, but which generate the same sequence $\mathbf{a}$.

In the case of sequences of period $q-1$, we need only determine one element $f_\alpha \in V_\alpha(\mathbf{a}) \in \mathbb{G}_n(\mathbf{a})$, and then determine the other elements of $\mathbb{G}_n(\mathbf{a})$ by composing f_α with nonlinear maps ϕ_γ^{-1} for each primitive element $\gamma \in \mathbb{F}_q^*$.

Remark 2. From the trace-representation of one filter generator (using a univariate polynomial $P(x) \in \mathbb{F}_q[x]/(x^q - x)$), it is much simpler to derive the trace-representation of the equivalent filter generators and then transform back to the ANF form. The univariate representation of the equivalent sequence generators are of the form $P(x^k)$, where all such polynomials have exactly the same weight and the equivalent functions are no more complicated in this sense.

The following proposition follows from lemma 3 and the discussion in this section.

Proposition 2. *Let* $\mathbf{s}_1, \mathbf{s}_2 \in \mathbb{F}_2^{q-1}$ *be sequences of period* $q-1$*, and assume there is* $\beta \in \mathbb{F}_q$ *a primitive element, such that* $V_\beta(\mathbf{s}_1) = V_\beta(\mathbf{s}_2)$*. Then* $\mathbb{G}_n(\mathbf{s}_1) = \mathbb{G}_n(\mathbf{s}_2)$*.*

6 Cryptanalytic Implications

If we restrict ourselves to keystream-sequences of period $q-1$, which is the common case for sequences generated by filter generators, then it follows from (4) that there are $2 \cdot |\mathbb{G}_n(\mathbf{s})|$ isomorphic filter generators generating the same keystream sequence, excluding affine equivalence. Thus, in order to assess the cryptographic properties of a filter generator, one should in theory check whether there exist in this class weak isomorphic ciphers with respect to some cryptographic property. In particular, it should be clear that any cryptographic property must be defined with respect to the weakest isomorphic cipher. This motivates a definition of the following type.

Definition 5. *Let* $\mathcal{P}$ *be a cryptographic measurement of a filter generator* $\mathcal{S}$*, which generates a sequence* $\mathbf{s}$*. Then the filter generator* $\mathcal{S}$ *is said to be* $\mathcal{P}$*-resistant only if there is no isomorphic filter generator* $\mathcal{S}'$ *with measurement* $\mathcal{P}' < \mathcal{P}$*.*

For example, consider a stream cipher $\mathcal{S}$ with a filter generator structure, which may employ a weak filter function that enables a successful algebraic attack. The results of the previous section imply that it is likely that there exists a cipher $\mathcal{S}'$ isomorphic to $\mathcal{S}$, that has a cryptographically much stronger Boolean function, and in turn may have been considered secure in that sense. An argument that supports this, while certainly not a proof, is given in Example 1 and in the next two subsections.

One can do the same type of argument with respect to any other filter generator, in that a randomly chosen isomorphic cipher may look more secure than a specifically designed instance, in the classical view of cryptanalysis. In cryptanalysis, it is clear that one would go for the weakest isomorphic cipher. It would in principle be possible to construct a trap-door function this way. However, such a direction would require further analysis, as such applications seem apparently inefficient in general.

Remark 3. Although out of the scope of this paper, as a general result, it would be interesting to divide $\mathbb{B}_n$ into classes of Boolean functions which are equivalent with respect to both nonlinear and linear equivalence. The main goal would be to

measure the amount of cryptographically strong Boolean functions. This could be achieved by dividing $\mathbb{B}_n$ into classes consisting of nonlinearly equivalent functions, together with the affine equivalences of those, and pick one representative from each such class. Such a class would be invariant regardless of the generator of $\mathbb{F}_q^*$. It should be noted that such a class would be much larger and general than the usual affine equivalences studied in literature, and would restrict the set of representatives of Boolean functions much further.

In the following section we discuss two properties of filter generators of cryptographic relevance, and how the results of this paper may be applied in the analysis of stream ciphers.

6.1 Algebraic Attacks

Algebraic attacks against stream ciphers were originally proposed by Courtois and Meier in [3]. The attack is a powerful technique against filter generators, and works by constructing systems of equations derived from the cipher operations, which can be solved using a choice of methods. Protection against algebraic attacks may for instance be reached by using filtering functions f of high degree, which neither f nor its complement $f + 1$ have low degree multiples. Algebraic degree and algebraic immunity are two properties of Boolean functions which are affine invariant. However, we have the following lemma when considering the equivalence $\mathbb{G}_n$.

Lemma 4. *The algebraic degree and algebraic immunity of a Boolean function f are not invariant with respect to $\mathbb{G}_n(\mathbf{s})$.*

This is clearly seen in Examples 1 and 2 (in the Appendix).

It is then for instance useful to have the following definition of algebraic immunity with respect to the equivalence $\mathbb{G}_n$.

Definition 6. *Let $f \in \mathbb{B}_n$ be a filter function used in a filter generator generating a sequence $\mathbf{s} \in \mathcal{L}_\alpha(f)$ of period $q - 1$, where we let $\mathbb{F}_2(\alpha) \simeq \mathbb{F}_q$. A more general algebraic immunity of a Boolean function f can be defined as*

$$\mathcal{GAI}(f) = min(\mathcal{AI}(f_\beta) \,|\, f_\beta \in V_\beta(\mathbf{s}),\ \textit{for all}\ V_\beta(\mathbf{s}) \in \mathbb{G}_n(\mathbf{s})).$$

However, it is not apparent whether the algebraic immunity of f_β is less than f or not in general. One could argue that if f contains less than n variables, then the equivalent functions will probably have higher algebraic immunity (since they probably involve all n variables). We consider this as a general open problem arising from our work.

6.2 Correlation Attacks

Correlation attack (see [11] and [7]) is another type of attack which has shown to be particularly successful against stream ciphers. A full treatment of the potential impact of our analysis on correlation attacks will be discussed on an forthcoming paper. Nevertheless, the purpose of this section is to show that:

1) *current analysis of distance from a nonlinear function to the space of affine (linear) functions is incomplete with respect to LFSR-based stream ciphers*;
2) *the notion of so-called weak feedback polynomials needs refinement.*

In order to address 1), we only need to point out the fact that there is not only one linear basis, but several. Assume that $\mathbb{F}_2(\alpha) \simeq \mathbb{F}_q$. If we let $g_{\alpha^k}(x) = \sum_{i=0}^{n_k}(x + \alpha^{k \cdot 2^i})$, where $n_k = \dim(\mathbb{F}_2(\alpha^k))$, it follows that

$$x^{q-1} - 1 = \prod_{k \in C(n)} g_{\alpha^k}(x),$$

where $C(n) \subset \{0, 1, 2, 3, \ldots, q-2\}$ denotes the coset-leaders modulo $q-1$.

In the following, for a polynomial $p(x) \in \mathbb{F}_2[x]$ dividing $x^q - x$, denote by

$$d_H(\, \Omega(p)\, ,\, \mathbf{a}\,) = (min(\, d_H(\, \mathbf{s},\, \mathbf{a})\,) \,|\, \mathbf{s} \in \Omega(p)),$$

the minimal distance between the vector space $\Omega(p)$ of sequences spanned by $p(x)$ and a sequence $\mathbf{a} \in \mathbb{F}_2^{q-1}$. Then we have the following definition of generalised correlations and distance to linear functions (linear subspaces).

Definition 7. *Let* $\mathbf{a} \in \mathbb{F}_2^{q-1}$. *Then define the minimal distance between* $\mathbf{a}$ *and a linear subspace by*

$$\mathcal{N}_1(\mathbf{a}) = min(\, d_H(\, \Omega(g_{\alpha^k})\, ,\, \mathbf{a}\,)\,) \,|\, 0 \leq k \leq q-2),$$

Assume that $\mathbf{a}$ is in $L_\alpha(f)$. Then, if for $l_2 \in \mathbb{B}_n$ we have that $d_H(\mathbf{a}, L_\alpha(l_1)) > d_H(\mathbf{a}, L_\beta(l_2))$ for all linear functions $l_1 \in \mathbb{B}_n$, it follows that a correlation attack is more successful on the equivalent function f_β.

Some correlation attacks (see for instance [4]) involve analysing LFSRs with low-weight feedback polynomials (or certain other nice properties). Such correlation analysis assume that the Boolean function models a *binary symmetric channel* (BSC) with certain correlation probability. Thus, it is sometimes possible to construct parity-check equations that relate the keystream to the underlying sequence-generator and allowing for instance one to mount a distinguishing attack. However, due to the fact that one may choose an equivalent filter generator with any desirable primitive polynomial (for instance a trinomial), it is clear that such analysis is not complete without taking into account the exact channel modelled by the Boolean function. If not, then this would mean that there always exists a cipher among the equivalent ciphers that is susceptible to correlation analysis, which is probably not true.

7 Conclusions and Future Research

Given a LFSR-based stream cipher $\mathcal{S}$ generating a sequence $\mathbf{s}$, we showed how to define an equivalence class $\mathbb{G}_n(\mathbf{s})$, consisting of all filter generators of length n that produce $\mathbf{s}$ as output (and in most cases of interest, of all filter generators equivalent to $\mathcal{S}$). In general, several properties of cryptographic relevance

are not invariant among the elements of $\mathbb{G}_n(\mathbf{s})$, and as a result it does not appear to make sense to conclude the security properties of a filter generator by, for instance, analysing the algebraic degree or algebraic immunity of the corresponding Boolean function, the properties such as the weight of the polynomial defining the LFSR, or the position of the registers that are tapped as input to the Boolean function. In particular, our analysis makes it clear that one cannot generally analyse the components of a stream cipher separately, as it is usual in practice. The natural object of analysis is the equivalence class $\mathbb{G}_n(\mathbf{s})$, and thus we believe that no analysis is complete without considering all of its elements.

Furthermore, we note that the idea presented here can be generalised into more complete equivalence classes. For example, instead of restricting oneself to the set of filter generators generating a particular sequence, one may instead define an equivalence with respect to the set of all possible combiner-generators generating a periodic sequence, in which cryptanalysis becomes much more fine-grained. We plan to explore this subject in more detail in a follow-up paper.

Acknowledgements

The work described in this paper was carried out while the first author was visiting Royal Holloway, University of London, supported by the Norwegian Research Council. The work has also been supported in part by the European Commission through the IST Programme under contract ICT-2007-216646 ECRYPT II.

References

1. Barkan, E., Biham, E.: How Many Ways Can You Write Rijndael? In: Zheng, Y. (ed.) ASIACRYPT 2002. LNCS, vol. 2501, pp. 160–175. Springer, Heidelberg (2002)
2. Cid, C., Murphy, S., Robshaw, M.J.B.: An Algebraic Framework for Cipher Embeddings. In: Smart, N.P. (ed.) Cryptography and Coding 2005. LNCS, vol. 3796, pp. 278–289. Springer, Heidelberg (2005)
3. Courtois, N., Meier, W.: Algebraic Attacks on Stream Ciphers with Linear Feedback. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 345–359. Springer, Heidelberg (2003)
4. Englund, H., Hell, M., Johansson, T.: Correlation attacks using a new class of weak feedback polynomials. In: Roy, B., Meier, W. (eds.) FSE 2004. LNCS, vol. 3017, pp. 127–142. Springer, Heidelberg (2004)
5. Golomb, S.W., Gong, G.: Signal Design for Good Correlation: For Wireless Communication, Cryptography, and Radar. Cambridge University Press, New York (2004)
6. Lidl, R., Niederreiter, H.: Introduction to Finite Fields and their Applications. Cambridge University Press, Cambridge (1994) (revised edition)
7. Meier, W., Staffelbach, O.: Fast correltaion attacks on stream ciphers (extended abstract). In: Günther, C.G. (ed.) EUROCRYPT 1988. LNCS, vol. 330, pp. 301–314. Springer, Heidelberg (1988)
8. Menezes, A.J., Van Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1996)

9. Murphy, S., Robshaw, M.J.B.: Essential Algebraic Structure Within the AES. In: Yung, M. (ed.) CRYPTO 2002. LNCS, vol. 2442, pp. 1–16. Springer, Heidelberg (2002)
10. Rønjom, S., Helleseth, T.: A new attack on the filter generator. IEEE Transactions on Information Theory 53(5), 1752–1758 (2007)
11. Siegenthaler, T.: Correlation-immunity of nonlinear combining functions for cryptographic applications. IEEE Transactions on Information Theory 30(5), 776–780 (1984)

A Appendix

The following example illustrates the lack of invariance of cryptographic properties of Boolean functions with respect to the equivalence classes $\mathbb{G}_5(\mathbf{s})$.

Example 2. Consider the binary sequence

$$\mathbf{s} = (10111111010001001100010101100 01),$$

of length 31. There are $\phi(31)/5 = 6$ primitive polynomials over $\mathbb{F}_2$ of degree 5. For each (distinct) generator β of the multiplicative group of $\mathbb{F}(\alpha)$, we compute a function f_β such that $\mathbf{s} \in \mathcal{L}_\beta(f_\beta)$, where we let $g_\alpha = x^5 + x^2 + 1$. The distinct nonzero coset-leaders modulo 31 are $K = \{1, 3, 5, 7, 11, 15\}$, and thus we may compute six functions $f_{\alpha_k}, k \in K$, where we let $\alpha_k = \alpha^k$ and pick one function f_{α_k} from each class $V_{\alpha_k} \in \mathbb{G}_5(\mathbf{s})$. The columns of the table below are ordered by the six functions $f_{\alpha_k} \in V_{\alpha_k}(\mathbf{s}) \in \mathbb{G}_5(\mathbf{s}), k \in K$:

	f_{α_1}	f_{α_3}	f_{α_5}	f_{α_7}	$f_{\alpha_{11}}$	$f_{\alpha_{15}}$
n	5	5	5	5	5	5
d	4	4	4	3	3	2
w_H	16	16	16	16	16	16
NL	10	10	10	8	12	8
AI	2	3	2	2	3	2
CI	0	0	0	1	0	1
PC	0	0	0	0	0	1
AB	16	16	16	16	8	32
SS	2432	2816	2816	3584	2048	8192

In the table above, w_H denotes the hamming weight of the functions, NL denotes *nonlinearity*, AI denotes *algebraic immunity*, CI denotes *correlation immunity*, PC denotes *propagation criterion of order 0*, AB denotes *absolute indicator* and SS denotes *sum-of-squares indicator*.

As one would expect, the weight of the truth-tables and the number of variables remains the same for each function. But notice that none of the other properties remain the same with respect to the transformation; and yet most

of these are properties that are invariant with respect to affine transformations. The functions are:

$$\begin{aligned}
f_{\alpha_1} &= x_0x_1x_2x_3 + x_0x_1x_2x_4 + x_0x_1x_3x_4 + x_1x_2x_3x_4 + x_0x_1x_2 + x_0x_1x_3 + \\
&\quad x_0x_2x_3 + x_1x_2x_3 + x_0x_1x_4 + x_2x_3x_4 + x_0x_2 + x_0 + x_1 \\
f_{\alpha_3} &= x_0x_1x_2x_3 + x_0x_1x_3x_4 + x_1x_2x_3x_4 + x_0x_1x_2 + x_0x_1x_4 + x_0x_3x_4 + \\
&\quad x_1x_3x_4 + x_2x_3x_4 + x_0x_1 + x_1x_3 + x_2x_4 + x_2 + x_3 \\
f_{\alpha_5} &= x_0x_1x_2x_4 + x_0x_2x_3x_4 + x_1x_2x_3x_4 + x_0x_1x_2 + x_0x_1x_3 + x_0x_2x_3 + \\
&\quad x_1x_2x_3 + x_0x_1x_4 + x_0x_3x_4 + x_0x_2 + x_0x_4 + x_1x_4 + x_2x_4 + x_0 + \\
&\quad x_1 + x_2 + x_3 + x_4 \\
f_{\alpha_7} &= x_0x_1x_3 + x_0x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2 + x_0x_3 + \\
&\quad x_0x_4 + x_1x_4 + x_3x_4 + x_0 + x_3 \\
f_{\alpha_{11}} &= x_0x_1x_2 + x_0x_2x_3 + x_1x_2x_3 + x_0x_1x_4 + x_1x_2x_4 + x_0x_1 + x_0x_2 + \\
&\quad x_1x_3 + x_0x_4 + x_2 \\
f_{\alpha_{15}} &= x_0x_1 + x_1x_2 + x_1x_3 + x_0x_4 + x_1x_4 + x_2x_4 + x_3x_4 + x_0 + x_1 + x_3
\end{aligned}$$

For instance, we now pick two of the above functions, say f_{α_1} and $f_{\alpha_{15}}$. If we let $\alpha_i = \alpha^i$, $X_1 \in \mathbb{F}_2(\alpha_1)^*$ and $X_{15} \in \mathbb{F}_2(\alpha_{15})^*$ and assume $X_1 = \alpha_1^{10}$, then we have that $X_{15} = X_1^{15} = (\alpha_1^{10})^{15} = \alpha_{15}^{10}$. Thus, if $S(X_1) = (1, 1, 1, 1, 0)$ denotes the initial state of an LFSR L_1 with generator polynomial $g_{\alpha_1}(x)$, then $S(X_{15}) = (1, 1, 0, 1, 1)$ denotes the initial state of the register L_2 with generator polynomial $g_{\alpha_{15}}(x)$. It then follows that

$$f_{\alpha_1}(S(X_1\alpha_1^t)) = f_{\alpha_{15}}(S(X_{15}\alpha_{15}^t)), t = 0, 1, 2, 3 \ldots,$$

and so the two different filter generators generate the same keystream-sequence

$$(00010011000101011000011011111101).$$

Thus, if one recovers the initial state of one cipher, it is a simple matter to recover the initial state of an isomorphic cipher. One would in this case for instance choose to attack the filter generator with the weakest function, say f_{15}.