# Open Problems Related to Algebraic Attacks on Stream Ciphers

Anne Canteaut[*]

INRIA - projet CODES,
B.P. 105 - 78153 Le Chesnay cedex, France
Anne.Canteaut@inria.fr

**Abstract.** The recently developed algebraic attacks apply to all keystream generators whose internal state is updated by a linear transition function, including LFSR-based generators. Here, we describe this type of attacks and we present some open problems related to their complexity. We also investigate the design criteria which may guarantee a high resistance to algebraic attacks for keystream generators based on a linear transition function.

## 1 Introduction

In an additive stream cipher, the ciphertext is obtained by adding bitwise the plaintext to a pseudo-random sequence called the keystream. The keystream generator is a finite state automaton whose initial internal state is derived from the secret key and from a public initial value by a key-loading algorithm. At each time unit, the keystream digit produced by the generator is obtained by applying a *filtering function* to the current internal state. The internal state is then updated by a *transition function.* Both filtering function and transition function must be chosen carefully in order to make the underlying cipher resistant to known-plaintext attacks. In particular, the filtering function must not leak too much information on the internal state and the transition function must guarantee that the sequence formed by the successive internal states has a high period.

Stream ciphers are mainly devoted to applications which require either an exceptional encryption rate or an extremely low implementation cost in hardware. Therefore, a linear transition function seems to be a relevant choice as soon as the filtering function breaks the inherent linearity. Amongst all possible linear transition functions, those based on LFSRs are very popular because they are appropriated for low-cost hardware implementations, produce sequences with good statistical properties and can be easily analyzed. LFSR-based generators have been extensively studied. It is known that the involved filtering function must

[*] This work was supported in part by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT. The information in this document reflects only the author's views, is provided as is and no warranty is given that the information is fit for any particular purpose. The user thereof uses the information at its sole risk and liability.

Ø. Ytrehus (Ed.): WWC 2005, LNCS 3969, pp. 120–134, 2006.

satisfy some well-defined criteria (such as a high nonlinearity, a high correlation-immunity order,...), and the designers of such generators now provide evidence that their ciphers cannot be broken by the classical attacks.

However, the recent progress in research related to algebraic attacks, introduced by Courtois and Meier [11], seems to threaten all keystream generators based on a linear transition function. In this context, it is important to determine whether such ciphers are still secure or not. Here, we investigate some related open problems, concerning the complexity of algebraic attacks (and of their variants) and concerning the design criteria of LFSR-based stream ciphers which guarantee a high resistance to these cryptanalytic techniques.

## 2 Basic Principle of Algebraic Attacks

Here, we focus on binary keystream generators based on a linear transition function, which can be described as follows. We denote by $\mathbf{x}_t$ the $n$-bit internal state of the generator at time $t$. The filtering function $f$ is first assumed to be a Boolean function of $n$ variables, i.e., at time $t$ the generator outputs only one bit, $s_t = f(\mathbf{x}_t)$. The transition function is supposed to be *linear* and is denoted by $L : \mathbf{F}_2^n \to \mathbf{F}_2^n$. Therefore, we have

$$s_t = f(L^t(\mathbf{x}_0)) \ ,$$

where $\mathbf{x}_0$ is the initial state. We only consider the case where both the filtering function and the transition function are publicly known, i.e., independent from the secret key. Two popular constructions known as nonlinear filter generators and combination generators fit the previous model.

The basic principle of algebraic attacks goes back to Shannon's work [26, Page 711]: these techniques consist in expressing the whole cipher as a large system of multivariate algebraic equations, which can be solved to recover the secret key. A major parameter which influences the complexity of such an attack is then the degree of the underlying algebraic system. When the transition is linear, any keystream bit can obviously be expressed as a function of degree $\deg(f)$ in the initial state bits. Therefore, it is known for a long time that the filtering function involved in such a stream cipher must have a high degree.

However, as pointed out by Courtois and Meier [11], the keystream generator may be vulnerable to algebraic attacks even if the degree of the algebraic function is high. Actually, the attack applies as soon as there exist relations of low degree between the output and the inputs of the filtering function $f$. Such relations correspond to low degree multiples of $f$, i.e., to relations $g(x)f(x) = h(x)$ for some $g$ where $h$ has a low degree. But, it was proved in [21, 24] that, in the case of algebraic attacks over $\mathbf{F}_2$, the existence of any such relation is equivalent to the existence of a low degree *annihilator* of $f$ or of $(1 + f)$, in the sense of Definition 1. Indeed, if $g(x)f(x) = h(x)$ with $\deg(h) \leq d$, we obtain, by multiplying this equation by $f(x)$, that

$$g(x) \left[f(x)\right]^2 = h(x)f(x) = g(x)f(x) = h(x) \ ,$$

leading to $h(x) \left[1 + f(x)\right] = 0$.

**Definition 1.** *Let $f$ be a Boolean function of $n$ variables. The* annihilator ideal *of $f$, denoted by $AN(f)$, is the set of all Boolean functions $g$ of $n$ variables such that*

$$g(x)f(x) = 0, \ \ \forall x \in \mathbf{F}_2^n \ .$$

*Moreover, for any degree $d$, we denote by $AN_d(f)$ the set of all annihilators of $f$ with degree at most $d$:*

$$AN_d(f) = \{g \in AN(f), \ \deg(g) \leq d\} \ .$$

Since the keystream bit at time $t$ is defined by $s_t = f \circ L^t(\mathbf{x_0})$, we deduce that:

- if $s_t = 1$, any function $g$ in $AN(f)$ leads to $g \circ L^t(\mathbf{x_0}) = 0$;
- if $s_t = 0$, any function $h$ in $AN(1 + f)$ leads to $h \circ L^t(\mathbf{x_0}) = 0$.

Therefore, if we collect the relations associated to all functions of degree at most $d$ in $AN(f) \cup AN(f + 1)$ for $N$ known keystream bits, we obtain a system of equations of degree $d$ depending on $n$ variables, $x_1, \ldots, x_n$, which correspond to the bits of the initial state:

$$\begin{cases} g \circ L^t(x_1, \ldots, x_n) \ \forall g \in AN_d(f), & \forall\, 0 \leq t < N \text{ such that } s_t = 1 \\ h \circ L^t(x_1, \ldots, x_n) \ \forall h \in AN_d(1 + f), & \forall\, 0 \leq t < N \text{ such that } s_t = 0 \end{cases} \quad (1)$$

The $n$-bit initial state can then be recovered by solving this multivariate polynomial system.

## 3 Complexity of Algebraic Attacks

Solving a multivariate polynomial system such as (1) is a typical problem studied in computer algebra. In order to get a rough estimate of the complexity of algebraic attacks for determining the suitable parameters for the keystream generator, we only focus on the simplest technique, called *linearization.* It consists in identifying the system with a linear system of $\sum_{i=1}^{d} \binom{n}{i}$ variables, where each product of $i$ bits of the initial state ($1 \leq i \leq d$) is seen as a new variable. The entire initial state is then recovered by a Gaussian reduction (or by more sophisticated techniques) whose time complexity is roughly

$$\left( \sum_{i=1}^{d} \binom{n}{i} \right)^{\omega} \simeq n^{\omega d} \ ,$$

where $\omega$ is the exponent of the matrix inversion algorithm, i.e., $\omega \simeq 2.37$ [9].

However, the previous estimation of the attack complexity is based on two hypotheses. It is first assumed that almost all monomials of degree $d$ appear in System (1). This clearly corresponds to the worst situation for the attacker, but we can wonder whether some weak choices for the transition function $L$ and for the filtering function $f$ can provide a system involving a small proportion of all possible monomials only, leading to a faster attack.

**Open problem 1.** *Determine the number of monomials in $x_1, \dots, x_n$ involved in System (1), depending on the choice of $L$ and $f$.*

A probably much stronger assumption in the usual complexity estimation is that the system can always be solved: it is usually supposed that the knowledge of

$$N \simeq \frac{2n^d}{d!\,(\dim A_d(f) + \dim A_d(1+f))}$$

keystream bits lead to a system with $\sum_{i=1}^{d} \binom{n}{i}$ linearly independent equations. It then raises the following open issue.

**Open problem 2.** *Determine the rank of System (1) depending on the choice of functions $L$ and $f$.*

Obviously, this question has an influence on the number of keystream bits required for the attack. But, a more crucial point is that the attack using equations of degree $d$ may be infeasible even if a huge keystream segment is available. This situation occurs when the system generated by $N$ keystream bits is underdetermined for any value of $N$. A natural related question is to determine whether the equations corresponding to a given annihilator $g$ are different for all keystream bits, i.e., whether there exists some $T$ less than the period of $\{L^t, t \geq 0\}$ such that $g \circ L^T(x) = g(x)$ for all $x \in \mathbf{F}_2^n$. It is clear that such an integer $T$ divides the period of $\{L^t, t \geq 0\}$. This observation leads to the following result when $L$ corresponds to the next-state function of an LFSR.

**Proposition 1.** *Let $L$ be the next-state function of an LFSR of length $n$ with primitive feedback polynomial. Let $g$ be a Boolean function of $n$ variables. If $2^n - 1$ is a prime, then all functions $g \circ L^t$, for $0 \leq t \leq 2^n - 1$, are distinct.*

But, when $(2^n - 1)$ is not a prime, there always exist filtering functions $f$ such that some of their annihilators $g \in AN(f)$, $g \neq 0$, lead to a sequence $\{g \circ L^t,\ 0 \leq t \leq 2^n - 1\}$ with a small period, as pointed out in the following toy example.

*Example 1.* Let us consider the LFSR of length 4 with primitive feedback polynomial $P(x) = x^4 + x + 1$ and the 4-variable filtering function $f$ defined by

$$f(x_1, \dots, x_4) = x_3 + x_4 + x_1x_2 + x_2x_3 + x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 \ .$$

Then, the function $g(x_1, \dots, x_4) = 1 + x_2 + x_3 + x_4 + x_2x_4 + x_3x_4$ belongs to $AN(f)$ and it satisfies

$$g \circ L^t(x_1, \dots, x_4) = g \circ L^{t \bmod 5}(x_1, \dots, x_4)$$

for all $t$. Actually, when $\mathbf{F}_2^4$ is identified with the finite field with 16 elements defined by the primitive polynomial $P$, we have $g(x) = g(x\alpha^5)$, where $\alpha$ is a root of $P$.

However, when a function $g$ in $AN(f)$ has such a strong periodic structure, this also holds for the filtering function, implying that the keystream can be easily distinguished from a random sequence.

**Proposition 2.** *Let $f$ be a Boolean function of $n$ variables and let $g$ be a nonzero function in $AN(f) \cup AN(1+f)$. If $g \circ L^T = g$ for some integer $T$, then there exists $t_0 < T$ such that all keystream bits $s_{t_0+iT}, i \geq 0$ are equal for at least one initial state. Moreover, if $L$ corresponds to the next-state function of an LFSR with primitive feedback polynomial, then all $s_{t_0+iT}, i \geq 0$ are equal for some $t_0 < T$ for all nonzero initial states when* $\deg(g) \neq n$.

*Proof.* Since $g$ is not the zero function, there exists some $a \in \mathbf{F}_2^n$ such that $g(a) = 1$, implying $g \circ L^{iT}(a) = 1$ for all $i \geq 0$. Because $g$ belongs to $AN(f)$ (resp. $AN(1+f)$), we deduce that $f$ (resp. $(1+f)$) vanishes at points $L^{iT}(a)$, for all $i \geq 0$. Therefore, the keystream generated from initial state $\mathbf{x}_0$ is such that $s_{t_0+iT}, i \geq 0$ are equal for some $t_0 < T$ as soon as an internal state $a$ with $g(a) = 1$ can be reached from $\mathbf{x}_0$. For an LFSR with maximum period, all internal states are generated for each nonzero $\mathbf{x}_0$, except the all-zero state. Thus, the property holds unless $g$ is the function of degree $n$ which vanishes at all points except 0.

However, the previous propositions only investigate the possibility that all equations derived from a given annihilator may be equal. The question of their linear dependency is still open. We can nevertheless conjecture from the previous discussion that, if the rank of the system involved in an algebraic attack highly differs from the rank of a random system, the corresponding keystream generator is probably vulnerable to a distinguishing attack.

If we assume that System (1) behaves like a random system with respect to both previously discussed properties, it clearly appears that the relevant parameter in the context of algebraic attacks against such stream ciphers is the so-called *algebraic immunity* of the filtering function.

**Definition 2.** *The* algebraic immunity *of a Boolean function $f$, denoted by $AI(f)$, is the lowest degree achieved by a nonzero function in $AN(f) \cup AN(1+f)$.*

It is worth noticing that the previous definition may be inappropriate when we consider algebraic attacks against other families of ciphers, for instance against block ciphers or combiners with memory. In such cases, the annihilator ideals of $f$ and of $(1+f)$ may play very different roles [3].

In our case, the time-complexity of algebraic attacks based on linearization is roughly

$$\mathcal{O}\left(n^{\omega AI(f)}\right) \text{ where } \omega \simeq 2.37$$

and the associated data-complexity, i.e., the required number of keystream bits, is $\mathcal{O}\left(n^{AI(f)}\right)$, but it is probably reduced when the number of functions of degree $AI(f)$ in $AN(f) \cup AN(1+f)$ increases. Thus, we can derive from this approximation a lower bound on the algebraic immunity of the filtering function which must be satisfied in order to resist algebraic attacks. If we suppose that the size of the internal state is minimal with respect to key-size $k$, i.e., that $n = 2k$ (it is known that the size of the internal state must be at least twice the key size in order to resist time-memory-data trade-off attacks), the complexity of the attack is greater than the complexity of an exhaustive search on the key when

$$AI(f) \geq 0.42 \left[\frac{k}{1+\log_2 k}\right] .$$

For instance, in a filter generator with a 128-bit key and a 256-bit internal state, the algebraic immunity of the filtering function must be at least 7.

But, the secure minimum value for the algebraic immunity is probably higher since more efficient techniques than linearization can be used for solving the algebraic system. Actually, this problem has been extensively studied in computer algebra and it is well-known that some methods based on Gröbner basis algorithms efficiently apply. The most recent and powerful algorithms, F4 and F5, are due to Faugère [19, 27, 20]. It was recently proved [18, 5] that F4 is more efficient than the extended linearization algorithm (XL) proposed by Courtois, Klimov, Patarin and Shamir [12]; XL actually computes a Gröbner basis in the particular context of algebraic attacks. And Algorithm F5 is strictly more efficient than all previous ones. Another technique, called XSL, has also been presented by Courtois and Pieprzyk [14] but its complexity and its implementation feasibility are still controversial.

Some recent results on the complexities of F4 and F5 can be found in [6, 7]. However, it is worth noticing that all these results only hold in the so-called semi-regular case. Therefore, the major problem is to determine whether the system involved in algebraic attacks behaves like a random system or not with respect to the previously mentioned algorithms. We would like to emphasize that it does not make sense to use some complexity results for the semi-regular case if we do not have any hint on the behaviour of the system. For instance, the public challenge on the asymmetric cryptosystem Hidden Field Equations (HFE) was broken by Faugère with F5 whereas the attack was infeasible according to its complexity in the generic case [22].

**Open problem 3.** *Does System (1) behave like a semi-regular system in the sense of [6]?*

## 4 Algebraic Immunity of Filtering Functions

Obviously, the algebraic immunity of the filtering function highly influences the complexity of the attack even if the estimation of the time complexity for solving the underlying system is still an open problem.

### 4.1 General Properties of the Algebraic Immunity

The set $AN(f)$ of all annihilating functions of $f$ is obviously an ideal in the ring of all Boolean functions, and it is generated by $(1+f)$. It consists of the $2^{2^n - wt(f)}$ functions of $n$ variables which vanish on the support of $f$, i.e., on all $x$ such that $f(x) = 1$, where $wt(f)$ denotes the size of the support of $f$. The number of functions of degree at most $d$ in $AN(f)$ is equal to $2^\kappa$ where $\kappa$ is the dimension of the kernel of the matrix obtained by restricting the Reed-Muller code of length $2^n$ and order $d$ to the support of $f$. In other words, the rows of this matrix correspond to the evaluations of the monomials of degree at most $d$ on $\{x, f(x) = 1\}$. Since this matrix has $\sum_{i=0}^{d} \binom{n}{i}$ rows and $wt(f)$ columns, its kernel is non-trivial when

$$\sum_{i=0}^{d} \binom{n}{i} > wt(f) \ .$$

Similarly, $AN(1+f)$ contains some functions of degree $d$ or less if

$$\sum_{i=0}^{d} \binom{n}{i} > 2^n - wt(f) \ .$$

Thus, as pointed out in [15], the algebraic immunity of an $n$-variable function is related to its Hamming weight. Most notably, for odd $n$, only balanced functions can have optimal algebraic immunity. A trivial corollary is also that, for any $n$-variable Boolean function, we have $AI(f) \leq \lceil n/2 \rceil$.

Another interesting property is that the highest possible algebraic immunity for a function is related to the number of its 0-linear structures. Let $\mathcal{S}_0(f)$ be the set of all 0-linear structures for $f$, i.e., $\mathcal{S}_0(f) = \{a \in \mathbf{F}_2^n, f(x+a) = f(x), \ \forall x\}$. Then,

$$AI(f) \leq \left\lceil \frac{n - \dim(\mathcal{S}_0(f))}{2} \right\rceil \ .$$

This bound is important for instance in the case of filtered LFSRs, since the filtering function usually depends only on a small subset of the internal state bits. We deduce from the previous discussion that if an $m$-variable Boolean function is used for filtering the $n$-bit internal state of the generator, the complexity of the algebraic attack will be at most $n^{\frac{\omega m}{2}}$. Therefore, the cipher resists algebraic attacks only if the number $m$ of variables of the filtering function satisfies

$$m \geq 0.84 \left[ \frac{k}{1 + \log_2(k)} \right] \ ,$$

where $k$ is the key-size and where the initial state is supposed to be twice longer than the key. For instance, a filter generator with a 128-bit key and a 256-bit internal state must use a filtering function of at least 16 variables. Here again, the secure number of variables is probably higher than the previous bound which is based on the complexity of linearization.

### 4.2 Algebraic Immunity of Random Balanced Functions

For 5-variable functions, it is possible to compute the algebraic immunity of all Boolean functions using the classification due to Berlekamp and Welch (because algebraic immunity is invariant under composition by a linear permutation). We here focus on balanced functions because they are the only ones that may have optimal algebraic immunity for $n$ odd. We can compute the algebraic immunity of all $601,080,390$ balanced functions of 5 variables:

Another interesting quantity is the number of linearly independent annihilators of degree at most 2 for all balanced functions of 5 variables:

| $AI(f)$ | 1 | 2 | 3 |
|---|---|---|---|
| nb. of balanced $f$ | 62 | 403,315,208 | 197,765,120 |
| proportion of balanced $f$ | $10^{-7}$ | 0.671 | 0.329 |

| $\dim(AN_2(f))$ | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| proportion of balanced $f$ | 0.329 | 0.574 | 0.094 | 0.002 | $2 \cdot 10^{-5}$ | $10^{-7}$ |

An important observation is that both sets $AN_2(f)$ and $AN_2(1+f)$ have the same dimension for all balanced functions except for one function and its complement (up to linear equivalence). This raises the following open problem.

**Open problem 4.** *For balanced Boolean functions $f$, is there a general relationship between $AN(f)$ and $AN(1+f)$?*

Similar simulations can be performed as far as the functions of $n$ variables are classified into equivalence classes under composition by a linear permutation. But, such a classification only exist for $n = 6$ and for cubic functions up to 8 variables.

Even if some well-known constructions of cryptographic Boolean functions have been proved to have a low algebraic immunity, probabilistic arguments tend to show that the proportion of balanced functions with low algebraic immunity is very small. It has been proved in [24] that the probability that a balanced function of $n$ variables has algebraic immunity less than $0.22n$ tends to zero when $n$ tends to infinity. An upper bound on the probability that a balanced function has an annihilator of degree less than $d$ is also given. This bound involves a part of the weight enumerator of $RM(d, n)$ and any new information on its complete weight distribution can clearly improve the result. However, both following problems are still open.

**Open problem 5.** *Determine the average value of the algebraic immunity for a balanced function of $n$ variables.*

**Open problem 6.** *Determine the proportion of balanced Boolean functions of $n$ variables with optimal algebraic immunity.*

### 4.3 Boolean Functions with Optimal Algebraic Immunity

A first relationship between the annihilators of $f$ and of $1 + f$ can be exhibited for functions with optimal algebraic immunity. Actually, all annihilators of a balanced $n$-variable function $f$ have maximal degree $\lfloor \frac{n+1}{2} \rfloor$ if and only if the support of $f$ corresponds to a subset of $2^{n-1}$ columns of the Reed-Muller code of length $2^n$ and order $\lfloor \frac{n-1}{2} \rfloor$ with maximal rank. When $n$ is odd, such a set is an information set for the Reed-Muller code of order $\frac{n-1}{2}$ which has dimension $2^{n-1}$. Then, a relationship between $\deg(AN(f))$ and $\deg(AN(1 + f))$ can be derived from the fact that this code is a self-dual code.

**Proposition 3.** *Let $\mathcal{C}$ be a linear self-dual code. If $I$ is an information set for $\mathcal{C}$, then its complement is an information set too.*

*Proof.* Let $I$ be an information set for $\mathcal{C}$. Then, there exists a generator matrix for $\mathcal{C}$ which can be decomposed into $G = (Id, M)_I$ where the first part corresponds to the positions in $I$. Let us now assume that the complement of $I$ is not an information set for $\mathcal{C}$. This means that there exists a nonzero codeword of the form $c = (c', 0)_I$ in $\mathcal{C}$. Since $\mathcal{C}$ is self-dual, $c$ belongs to the dual code. Therefore, $Gc = 0$, implying that some columns of the identity matrix sum up to zero, a contradiction.

We can immediately derive the following result.

**Theorem 1.** *Let $n$ be an odd integer and $f$ be a balanced Boolean function of $n$ variables. Then, $f$ has optimal algebraic immunity $\frac{n+1}{2}$ if and only if $AN(f)$ does not contain any nonzero function of degree strictly less than $\frac{n+1}{2}$.*

A few classes of Boolean functions with optimal algebraic immunity have been recently exhibited. An iterative construction which provides an infinite family of balanced Boolean functions with optimal algebraic immunity is presented in [16]. Another example of functions with optimal algebraic immunity is the majority symmetric function depending on an odd number of variables, i.e., the function which outputs 1 if and only if the Hamming weight of its input vector is greater than or equal to $\frac{n+1}{2}$. This property was first proved in [23, Theorem 1] in terms of information sets for the self-dual Reed-Muller code, and it is also mentioned in [17].

### 4.4 Algebraic Immunity and Other Cryptographic Criteria

Besides the Hamming weight of the function, its nonlinearity is also related to its algebraic immunity [15]. It can be proved that, for any linear function $\varphi$, the algebraic immunity of $f + \varphi$ is at most $AI(f) + 1$. Therefore, any function $f$ of $n$ variables with algebraic immunity at least $d$ satisfies

$$\mathcal{NL}(f) \geq \sum_{i=0}^{d-2} \binom{n}{i} .$$

It follows that any function with optimal algebraic immunity has a high nonlinearity, more precisely

$$\mathcal{NL}(f) \geq \begin{cases} 2^{n-1} - \binom{n}{\frac{n-1}{2}} & \text{if } n \text{ is odd} \\ 2^{n-1} - \frac{1}{2}\binom{n}{\frac{n}{2}} - \binom{n}{\frac{n}{2}-1} & \text{if } n \text{ is even} \end{cases}$$

A high nonlinearity and a high algebraic immunity are then compatible criteria. Another important consequence is that the nonlinearity of a function may be a sufficient criterion to decide whether it has low algebraic immunity (but the converse is not true).

Another cryptographic property that implies that a function does not have a maximal algebraic immunity is the notion of *normality.* A function is said to be $k$-normal (resp. $k$-weakly normal) if there exists an affine subspace of dimension $k$ on which the function is constant (resp. affine). Since the minimum weight codewords of $RM(r,n)$ are those whose support is an affine subspace of dimension $n-r$, we deduce that any $k$-normal function $f$ of $n$ variables has algebraic immunity at most $n-k$. Similarly, any $k$-weakly normal function has algebraic immunity at most $n-k+1$. Non-normal (and non-weakly normal) functions may be good candidates if we want to construct functions with optimal nonlinearity.

The existence of links between algebraic immunity and other cryptographic criteria remains unknown. For instance, the relation between the distance of a function to all low-degree functions (i.e., its distance to $RM(d,n)$) and its algebraic immunity is still unclear. Correlation-immunity does not seem to be a priori incompatible with optimal algebraic immunity: there exists a 1-resilient function of 5 variables with optimal algebraic immunity. However, the link with all known criteria must be investigated further.

### 4.5 Algebraic Immunity of Known Constructions

Some bounds have been established on the algebraic immunity of the cryptographic functions obtained by applying classical constructions. First, the algebraic immunity of a function can be derived from the algebraic immunities of its restrictions to a given hyperplane and to its complement [15]. For instance, if

$$f(x_1,\dots,x_n) = (1+x_n)f_1(x_1,\dots,x_{n-1}) + x_n f_2(x_1,\dots,x_{n-1}) \ ,$$

we have:

– if $AI(f_1) \neq AI(f_2)$, then $AI(f) = \min(AI(f_1), AI(f_2)) + 1$;
– if $AI(f_1) = AI(f_2)$, then $AI(f) \in \{AI(f_1), AI(f_1)+1\}$.

Therefore, it is obvious how to construct a function of $2t$ variables with optimal algebraic immunity from two functions of $(2t-1)$ variables with respective algebraic immunities equal to $t$ and to $(t-1)$. But, constructing a function of $(2t+1)$ variables with optimal algebraic immunity from two functions of $2t$ variables is much more difficult since both restrictions must have optimal algebraic immunity and they must also satisfy some additional conditions.

Some bounds on the algebraic immunities of some classical constructions, such as the Maiorana-McFarland family, can be found in [24, 15, 25].

### 4.6 Computing the Algebraic Immunity of a Boolean Function

The basic algorithm for computing the algebraic immunity of an $n$-variable function consists in performing a Gaussian elimination on the generator matrix of the punctured $RM(\lfloor \frac{n-1}{2} \rfloor, n)$ restricted to the support of $f$. This matrix has $wt(f)$ columns and $k(\lfloor \frac{n-1}{2} \rfloor, n) = \sum_{i=0}^{\lfloor \frac{n-1}{2} \rfloor} \binom{n}{i}$ rows. Therefore, the algorithm requires $k^2(\lfloor \frac{n-1}{2} \rfloor, n) wt(f)$ operations, which is close to $2^{3n-3}$ when $f$ is balanced. As noted in [24], the complexity can be significantly reduced if we only want to check whether a function has annihilators of small degree $d$, since we do not need to consider all positions in the support of $f$. Indeed, considering a number of columns which is only slightly higher that the code dimension $k(d,n)$

is usually sufficient for proving that a function does not admit any annihilator of degree $d$. A technique for reducing the size of the matrix over which the Gaussian elimination is performed is presented in [24]. The idea is that the elements in the support of $f$ with low Hamming weight provide simple equations that can be removed from the matrix by a substitution step. However, due to the lack of simulation results, it is very hard to evaluate the time complexity of the substitution step in practice.

Gröbner bases algorithms such as F5 provide other techniques for computing the size of the annihilator ideal. But they need to be compared with the basic techniques in this particular context.

## 5 Resistance to Fast Algebraic Attacks

At CRYPTO 2003, Courtois presented some important improvements on algebraic attacks, called *fast algebraic attacks* [10]. The refinement first relies on the existence of some low degree relations between the bits of the initial state and not only one but several consecutive keystream bits. In other words, the attacker wants to find some low degree relations $g$ between the inputs and outputs of

$$\begin{array}{rl} F_m\colon \mathbf{F}_2^n & \rightarrow \mathbf{F}_2^m \\ x & \mapsto (f(x), f(L(x)), \ldots, f(L^{m-1}(x)) \end{array}$$

where $L$ is the linear transition function. This function is very similar to the so-called *augmented function* defined in [1]. The fact that the augmented function may be much weaker than the filtering function, i.e., than $F_0$ with the previous notation, has been pointed out by Anderson [1] in the context of correlation attacks. However, finding the low degree relations between the $n$ inputs and $m$ outputs of $F_m$ becomes infeasible when $m$ increases. The direct algorithm used for a function $S$ with $n$ inputs and $m$ outputs consists in finding the low degree annihilators for the characteristic function $\Phi_S$ of $S$, which is the Boolean function of $(n+m)$ variables defined by

$$\Phi_S(x_1, \ldots, x_n, y_1, \ldots, y_m) = 1 \text{ if and only if } y_i = S_i(x_1, \ldots, x_n), \ \forall i \ .$$

Due to its high complexity, it can only be used for small values of $m$. For instance, if we consider a Boolean function of 20 variables, it may have algebraic immunity 10. But, there always exist relations of degree at most 7 involving 4 consecutive keystream bits together. The problem is that determining whether relations of degree less than or equal to 6 exist in this case requires the computation of the kernel of a matrix of 120 GBytes. And even checking whether relations of degree 3 exist involves a 2.7 GByte-matrix. Mounting algebraic attacks based on the augmented function is then related to the following problem.

**Open problem 7.** *Find an algorithm which determines the low-degree relations for the augmented function.*

More generally, we can wonder whether the particular form of the augmented function has an influence on the degree of the annihilator ideal of its characteristic function. For instance, the existence of a general relationship between the algebraic immunity of a Boolean function and the algebraic immunity of the associated augmented function is still unclear. The fact that the augmented function is a very special case of multi-output functions may lead to new theoretical results or to dedicated algorithms in that case. For instance, a very particular property of the augmented function is that all its Boolean components are linearly equivalent. This raises the following open question, which is clearly related to algebraic attacks against block ciphers which use power functions as S-Boxes, like the AES.

**Open problem 8.** *Does the linear equivalence between all output components of a multi-output function influence its algebraic immunity?*

Since the computation of low degree relations involving several keystream bits is usually infeasible, Courtois proposed to focus on particular subclasses of relations that can be obtained much faster. The relations considered in the attack are given by linear combinations of relations of the form

$$g(x_0, \ldots, x_{\ell-1}, s_t, \ldots, s_{t+m})$$

where the terms of highest degree do not involve any keystream bits. Then, an additional precomputation step consists in determining the linear combinations of the previous relations which cancel out the highest degree monomials. Some algorithms for this step have been proposed in [10, 2]. This technique helps to decrease the degree of the relations used in the attack for different practical examples. But, here again, we do not have any theoretical result connecting the algebraic immunity of the function and the existence of such low degree linear combinations.

## 6 Using More Sophisticated Filtering Functions

Many stream ciphers do not use a simple Boolean filtering function; they prefer more sophisticated mappings in order to render the attacks more difficult or in order to increase the throughput of the generator.

*Multi-output Boolean functions.* A basic technique for increasing the speed of the generator consists in using a filtering function with several outputs. Such functions are called vectorial Boolean functions, or S-boxes by analogy with block ciphers. But, as pointed out in [28], the resistance of the generator to fast correlation attacks usually decreases with the number of output bits of the function. For a single output function, the attack exploits the fact that the output may be approximated by an affine function of the input variables. But, for a function $S$ with $m$ outputs, the attacker can apply any Boolean function $g$ of $m$ variables to the output vector $(y_1, \ldots, y_m)$ and he or she can perform the attack on the resulting sequence $z = g(y_1, \dots, y_m)$. Therefore, the relevant parameter is not the nonlinearity of the vectorial function, which is the lowest Hamming distance between any linear combination of the components of $S$ and the affine functions, but the so-called *unrestricted nonlinearity* [8], which is the lowest distance between any function $g \circ S$ and the affine functions, where $g$ varies in the set of all nonzero Boolean functions of $m$ variables.

For similar reasons, the algebraic immunity of a vectorial function tends to decrease with the number of output bits. For an S-box with $n$ inputs and $m$ outputs, there exists a relation of degree at most $d$ in the input variables (and of any degree in the output variables) if

$$\sum_{i=0}^{d} \binom{n}{i} > 2^{n-m} .$$

A particular case of generators based on multi-output Boolean functions are the word-oriented ciphers. In order to increase the performance of software implementations, many ciphers use LFSRs over an extension field $\mathbf{F}_{2^m}$ and the associated filtering function is usually a mapping from $\mathbf{F}_{2^m}^n$ into $\mathbf{F}_{2^m}$. This technique is used in many recent stream ciphers, e.g. in SNOW 2.0. The associated filtering function can obviously be seen as a vectorial Boolean function with $mn$ inputs and $m$ outputs. Consequently, all results previously mentioned apply, but the major open issue here is to determine whether word-oriented attacks can be mounted which exploit the particular structure of the function defined as a polynomial over $\mathbf{F}_{2^m}$.

*Functions with memory.* In some keystream generators, the filtering function is replaced by a finite automaton with some memory bits. An example is the $E_0$ keystream generator used in the Bluetooth wireless LAN system, which uses a combining function with 4 inputs and 4 memory bits. However, (fast) algebraic attacks [4] can still be applied on such systems. Armknecht and Krause proved that, for any filtering function of $n$ variables with $M$ memory bits, there always exists a relation of degree at most $\lceil \frac{n(M+1)}{2} \rceil$ between $(M+1)$ consecutive output bits and the bits of the initial state, for a given initial assignment of the memory bits. Obviously, relations of lower degree may exist. For instance, the function used in $E_0$ provides a relation of degree 4 involving 4 consecutive output bits, which leads to an algebraic attack of running-time around $2^{67}$ [4]. General results on algebraic attacks against combiners with memory can be found in [3, 13].

The main open issue related to the use of such sophisticated functions is to improve the efficiency of the algorithms for computing their algebraic immunity for a large number of input variables. Another related open problem is to find some general constructions which guarantee a high resistance to all these attacks.

## Acknowledgements

Many thanks to Daniel Augot and Matthew Parker for their contributions to this work.

## References

1. R. J. Anderson. Searching for the optimum correlation attack. In *Fast Software Encryption - FSE'94*, volume 1008 of *Lecture Notes in Computer Science*, pages 137–143. Springer-Verlag, 1995.
2. F. Armknecht. Improving fast algebraic attacks. In *Fast Software Encryption - FSE 2004*, volume 3017 of *Lecture Notes in Computer Science*, pages 65–82. Springer-Verlag, 2004.
3. F. Armknecht. Algebraic attacks and annihilators. In *Proceedings of the Western European Workshop on Research in Cryptology (WEWoRC 2005)*, Lecture Notes in Informatics. Springer-Verlag, 2005. To appear.
4. F. Armknecht and M. Krause. Algebraic attacks on combiners with memory. In *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 162–176. Springer-Verlag, 2003.
5. G. Ars, J.-C. Faugère, H. Imai, M. Kawazoe, and M. Sugita. Comparison between XL and Gröbner basis algorithms. In *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *Lecture Notes in Computer Science*, pages 338–353. Springer-Verlag, 2004.
6. M. Bardet, J-C. Faugère, B. Salvy, and B-Y. Yang. On the complexity of Gröbner basis computation of semi-regular overdetermined algebraic equations. In *Proc. International Conference on Polynomial System Solving (ICPSS'2004)*, 2004.
7. M. Bardet, J-C. Faugère, B. Salvy, and B-Y. Yang. Asymptotic behaviour of the degree of regularity of semi-regular polynomial systems. In *MEGA 2005*, Porto Conte, Italy, May 2005.
8. C. Carlet and E. Prouff. On a new notion of nonlinearity relevant to multi-output pseudo-random generators. In *Selected Areas in Cryptography - SAC 2003*, volume 3006 of *Lecture Notes in Computer Science*, pages 291–305. Springer-Verlag, 2004.
9. D. Coppersmith and S. Winograd. Matrix multiplication via arithmetic programming. *Journal of Symbolic Computation*, (9):251–280, 1990.
10. N. Courtois. Fast algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*, pages 176–194. Springer-Verlag, 2003.
11. N. Courtois and W. Meier. Algebraic attacks on stream ciphers with linear feedback. In *Advances in Cryptology - EUROCRYPT 2003*, volume 2656 of *Lecture Notes in Computer Science*, pages 345–359. Springer-Verlag, 2003.
12. N. T. Courtois, A. Klimov, J. Patarin, and A. Shamir. Efficient algorithms for solving overdefined systems of multivariate polynomial equations. In *Advances in Cryptology - EUROCRYPT 2000*, volume 1807 of *Lecture Notes in Computer Science*, pages 392–407. Springer-Verlag, 2000.
13. N.T. Courtois. Algebraic attacks on combiners with memory and several outputs. In *ICISC 2004*, volume 3506 of *Lecture Notes in Computer Science*, pages 3–20. Springer-Verlag, 2005.
14. N.T. Courtois and J. Pieprzyk. Cryptanalysis of block ciphers with overdefined systems of equations. In *Advances in Cryptology - ASIACRYPT 2002*, volume 2502 of *Lecture Notes in Computer Science*, pages 267–287. Springer-Verlag, 2002.
15. D.K. Dalai, K.C. Gupta, and S. Maitra. Results on algebraic immunity for cryptographically significant Boolean functions. In *Progress in Cryptology - Indocrypt 2004*, volume 1880 of *Lecture Notes in Computer Science*, pages 92–106. Springer-Verlag, 2004.

16. D.K. Dalai, K.C. Gupta, and S. Maitra. Cryptographically significant Boolean functions: Construction and analysis in terms of algebraic immunity. In *Fast Software Encryption - FSE 2005*, volume 3357 of *Lecture Notes in Computer Science*, pages 98–111. Springer-Verlag, 2005.
17. D.K. Dalai, S. Sarkar, and S. Maitra. Balanced Boolean functions with maximum possible algebraic immunity. Preprint, April 2005.
18. C. Diem. The XL algorithm and a conjecture from commutative algebra. In *Advances in Cryptology - ASIACRYPT 2004*, volume 3329 of *Lecture Notes in Computer Science*, pages 323–337. Springer-Verlag, 2004.
19. J.-C. Faugère. A new efficient algorithm for computing Gröbner bases $(F_4)$. *Journal of Pure and Applied Algebra*, 139(1-3):61–88, 1999.
20. J.-C. Faugère. A new efficient algorithm for computing Gröbner bases without reduction to zero $(F_5)$. In *Proceedings of the 2002 international symposium on Symbolic and algebraic computation*. ACM, 2002.
21. J.-C. Faugère and G. Ars. An algebraic cryptanalysis of nonlinear filter generators using Gröbner bases. Technical Report 4739, INRIA, 2003. Available at `ftp://ftp.inria.fr/INRIA/publication/publi-pdf/RR/RR-4739.pdf`.
22. J.-C. Faugère and A. Joux. Algebraic cryptanalysis of hidden field equation (HFE) cryptosystems using Gröbner bases. In *Advances in Cryptology - CRYPTO 2003*, volume 2729 of *Lecture Notes in Computer Science*. Springer-Verlag, 2003.
23. J.D. Key, T.P. McDonough, and V.C. Mavron. Information sets and partial permutation decoding for codes from finite geometries. *Finite Fields and Their Applications*, 2005. To appear.
24. 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. Springer-Verlag, 2004.
25. E. Pasalic. On algebraic immunity of Maiorana-McFarland like functions and applications of algebraic attack. In *Proceedings of the ECRYPT Symmetric Key Encryption Workshop (SKEW)*, Aarhus, Danemark, May 2005.
26. C.E. Shannon. Communication theory of secrecy systems. *Bell system technical journal*, 28:656–715, 1949.
27. A. Steel. Allan Steel's Gröbner basis timings page, 2004. `http://magma.maths.usyd.edu.au/users/allan/gb/`.
28. M. Zhang and A. Chan. Maximum correlation analysis of nonlinear S-boxes in stream ciphers. In *Advances in Cryptology - CRYPTO 2000*, volume 1880 of *Lecture Notes in Computer Science*, pages 501–514. Springer-Verlag, 2000.