Vectorial Boolean Functions and Induced Algebraic Equations

Jovan Dj. Golić
Access Network and Terminals System Design
Telecom Italia Lab, Telecom Italia
Via Reiss Romoli 274, 10148 Turin, Italy
Email: jovan.golic@tilab.com

September 1, 2004

Abstract

A general mathematical framework behind algebraic cryptanalytic attacks is developed. The framework relates to finding algebraic equations induced by vectorial Boolean functions and, in particular, equations of low algebraic degree. The equations may involve only a subset of input variables and may or may not be conditioned on the values of output variables. In addition, the equations may have a special form interesting for the so-called fast algebraic attacks. A possible divide-and-conquer effect is pointed out and the notion of algebraic immunity order, naturally extending the notion of correlation immunity order, is introduced. An application of general results to stream ciphers known as combiners with or without memory, with possibly multiple outputs, is studied in particular detail. Special properties of combiners with finite input memory, such as nonlinear filter generators, are established. Finally, finding induced algebraic equations for divide-and-conquer algebraic attacks on combiners with or without memory is also considered.

Keywords: Vectorial Boolean functions, algebraic equations, algebraic attacks, combiners with memory

1 Introduction

Algebraic attacks are cryptanalytic attacks that reconstruct the secret key in cryptosystems by manipulating and solving the underlying algebraic equations. Of course, solving algebraic equations is generally a difficult problem, but it has been demonstrated that for particular ciphers, finding solutions faster than by the exhaustive search may sometimes be possible. This is especially the case if the algebraic equations have a relatively low algebraic degree when considered as multivariate polynomial equations. If the number of such equations is sufficiently large, then it may be possible to apply the well-known linearization algorithm, which consists of replacing the products of variables, that is, the monomials by new variables and of solving the resulting system of linear equations, in new variables, for example, by Gaussian elimination. It is shown in [12] that even if the number of low-degree equations is not large enough for the linearization algorithm to work, one can derive new equations of increased degree by multiplying the original equations by monomials of a limited degree so that, provided that the original system is overdefined, the new system may contain sufficiently many linearly independent equations for the linearization algorithm to work. This algorithm is known as the XL algorithm. It is demonstrated in [2] that the linearization and XL algorithms may even be applicable if the low-degree algebraic equations do not hold with certainty, but with probabilities sufficiently close to 1.

When the XL algorithm is applied to product block ciphers, it is shown in [3] that the complexity grows polynomially with the number of rounds, but, nevertheless, the algorithm is not so effective or is on the borderline to be effective, because of a large number of intermediate variables introduced and because of possible linear dependences among the derived equations.

In [4], it is shown that algebraic attacks may be effective against stream ciphers known as memoryless combiners and nonlinear filter generators, in which the output is produced by a Boolean function applied to the internal state sequence generated by an autonomous linear finite-state machine, for example, by a number of linear feedback shift registers (LFSR's). In this case, the degree of the algebraic equations is preserved when the inputs to the Boolean function are expressed as linear functions of the secret key. The main points of [4] are that a Boolean function of a given algebraic degree may induce algebraic equations of lower degrees and that for any Boolean function of k variables, there exists an induced algebraic equation of degree at most $\lceil k/2 \rceil$. As a consequence, if the number of such algebraic equations induced by known keystream bits is sufficiently large, then the secret key can be recovered by the linearization algorithm.

This result is extended to binary combiners with k inputs and l bits of memory in [1], where it is proven that there exists an induced algebraic equation over $l+1$ consecutive k-bit inputs of degree $\lceil k(l+1)/2 \rceil$, when conditioned on $r = l+1$ consecutive output bits. A generalization to combiners with memory and multiple outputs is given in [6], but the existence of non-trivial equations involving only the input variables is not guaranteed. This is due to the specific unconditional scenario, where the algebraic equations are allowed to involve the output variables, with no restriction regarding their algebraic degree.

The low-degree algebraic equations induced by r consecutive output bits can be found by solving the underlying system of linear equations, for example, by Gaussian elimination, with complexity $O(2^{3kr})$. Some complexity reductions are proposed in [11].

It may be interesting to point out the analogy between finding low-degree induced algebraic equations and finding linear correlations in combiners with memory, where the objective is to find linear equations, i.e., algebraic equations of degree 1 that are allowed to hold with probabilities different from one half. For example, see [10] and [7], whereas for combiners with multiple outputs, see [8].

In [5], it is shown that the algebraic attacks can be faster if a found low-degree equation in the input and output variables has a special form, that is, if the monomials with the highest degrees in input variables do not depend on output variables. As a result, these monomials can be eliminated by using the linear complexity properties of the corresponding stream cipher, so that the algebraic degree is thus effectively reduced. Some further complexity clarifications and reductions can be found in [9].

The first objective of this paper is to provide a unified mathematical framework for finding induced algebraic equations of low algebraic degree, for arbitrary vectorial Boolean functions in the conditional, unconditional, and constrained unconditional scenarios, where a subset of input variables may be required to be eliminated from the equations. The second objective is to generalize the notion of correlation immunity order [13] to that of alebraic immunity order, for any vectorial Boolean function, and to study its properties. The third objective is to apply the general results to combiners with or without memory and thus obtain some new results relating to the three scenarios considered, to combiners with finite input memory, to the influence of multiple outputs, and to the divide-and-conquer effect achievable. Sections 2 and 3 are devoted to the first two objectives, respectively, whereas the third objective is dealt with in Sections 4, 5, and 6. Very compact proofs of the basic mathematical results are provided in Section 2, whereas the proofs of the derived results from the remaining sections are left to the reader.

2 Induced Algebraic Equations

Consider a vectorial Boolean function $f : \{0,1\}^{n_x} \times \{0,1\}^{n_y} \to \{0,1\}^{n_z}$, denoted as $Z = f(X,Y)$. Let $n_z^* = \log|f_z|$, where $f_z = \{Z : (\exists\, X,Y)\, f(X,Y) = Z\}$ denotes the range of f. (Here and throughout, the logarithms are always taken to the base 2.) Further, let $f^{-1}(Z) = \{(X,Y) : f(X,Y) = Z\}$ and $f_x^{-1}(Z) = \{X : (\exists\, Y)\, f(X,Y) = Z\}$. Also, let $n_{x,z}^* = \log|f_{x,z}|$, where $f_{x,z} = \{(X,Z) : (\exists\, Y)\, f(X,Y) = Z\}$.

Our objective is to study algebraic equations induced by f, that is, by the equation $f(X,Y) = Z$. In the conditional scenario, Z is fixed and the algebraic equations involve only X, whereas in the unconditional scenario, Z is variable and the algebraic equations involve both X and Z. Accordingly, in both scenarios, Y is not allowed to be involved in the algebraic equations. An algebraic equation is specified by a (non-zero) Boolean function defined itself by a multivariate binary polynomial, that is, by an algebraic normal form, and this Boolean function is required to be equal to zero on a considered subset of values of the involved variables.

In general, for any subset $\mathcal{S} \subset \{0,1\}^n$, a non-trivial algebraic equation induced by $\mathcal{S}$ is an equation of the form $g(S) = 0$, $S \in \mathcal{S}$, where g is a non-zero multivariate polynomial in $S = (s_1, \cdots, s_n)$, that is, a non-zero Boolean function $\{0,1\}^n \to \{0,1\}$ specified by its algebraic normal form. We equivalently say that such an algebraic equation is induced by $\mathcal{S}$ or is satisfied on $\mathcal{S}$. The trivial algebraic equation is defined by the zero polynomial g. It follows that the set of all such g is a vector space, as it is closed under addition. In particular, we are interested in algebraic equations of low (algebraic) degree, which are defined by multivariate polynomials of degree at most d, where $1 \leq d \leq n$. The set of all such algebraic equations is also a vector space. Let $\mathcal{P}_n^d$ denote the set of all the multivariate polynomials in n variables of degree at most d. Any such polynomial can be characterized as a linear combination of all $\sum_{i=0}^{d} \binom{n}{i}$ (linearly independent) monomials of degree at most d. Let $\mathbf{P}_n^d(\mathcal{S})$ be an $|\mathcal{S}| \times \sum_{i=0}^{d} \binom{n}{i}$ matrix whose rows and columns are indexed by the vectors from $\mathcal{S}$ and by the monomials from $\mathcal{P}_n^d$, respectively, and whose entries are defined by evaluating the monomials on these vectors.

The following lemmas are essentially behind all the theorems to be derived in the sequel, which in fact correspond to particular subsets $\mathcal{S}$. As such, they are also behind the known results on combiners with or without memory from [4], [1], [5], and [6].

Lemma 1 *For any $\mathcal{S} \subset \{0,1\}^n$, there exists a non-trivial algebraic equation on $\mathcal{S}$ iff $|\mathcal{S}| < 2^n$.*

Proof By definition, a non-trivial algebraic equation on $\mathcal{S}$ is defined by a non-zero Boolean function $g(S)$, $S \in \{0,1\}^n$, such that $g(S) = 0$, $S \in \mathcal{S}$. Such a function exists iff the set $\{0,1\}^n \setminus \mathcal{S}$ is not empty. □

Lemma 2 *For any $\mathcal{S} \subset \{0,1\}^n$ and $g \in \mathcal{P}_n^d$, g defines a non-trivial algebraic equation on $\mathcal{S}$ iff the columns of $\mathbf{P}_n^d(\mathcal{S})$ corresponding to the monomials in g add up to zero.*

Proof As each $g \in \mathcal{P}_n^d$ is a linear combination $g = \sum_k c_k\, m_k$ of $\sum_{i=0}^d \binom{n}{i}$ different monomials m_k from $\mathcal{P}_n^d$, it follows that the equations $g(S) = 0$ and $\sum_k c_k\, m_k(S) = 0$, $S \in \mathcal{S}$, are equivalent. The claim then follows from the definition of the matrix $\mathbf{P}_n^d(\mathcal{S})$. □

Lemma 3 *For any $\mathcal{S} \subset \{0,1\}^n$, there exists a $g \in \mathcal{P}_n^d$ that defines a non-trivial algebraic equation on $\mathcal{S}$ if*

$$\sum_{i=0}^{d} \binom{n}{i} > |\mathcal{S}|. \tag{1}$$

Proof If (1) is true, then the number of columns exceeds the number of rows in $\mathbf{P}_n^d(\mathcal{S})$. As a consequence, the columns are linearly dependent and the claim then follows from Lemma 2. □

If the binary coefficients defining any $g \in \mathcal{P}_n^d$ as a linear combination of all $\sum_{i=0}^d \binom{n}{i}$ monomials of degree at most d are represented as a binary one-column matrix $\mathbf{C}$, then the vector space of all algebraic equations induced by $\mathcal{S}$ can be obtained by solving the system of linear equations $\mathbf{P}_n^d(\mathcal{S}) \cdot \mathbf{C} = \mathbf{0}$ in $\mathbf{C}$, for example, by Gaussian elimination with time complexity $O\left(|\mathcal{S}| \left(\sum_{i=0}^d \binom{n}{i}\right)^2\right)$ and space complexity $O\left(|\mathcal{S}| \sum_{i=0}^d \binom{n}{i}\right)$. The dimension of this vector space equals $\sum_{i=0}^d \binom{n}{i} - \operatorname{rank} \mathbf{P}_n^d(\mathcal{S})$.

2.1 Conditional Algebraic Equations

In the conditional scenario, given a vectorial Boolean function $Z = f(X, Y)$, the objective is to find algebraic equations over X when Z is assumed to be fixed and known. Variables in Y should thus be eliminated. This scenario is particularly interesting for the cryptanalysis of stream ciphers in the known keystream sequence scenario, provided that f is used as an output function producing a binary keystream sequence from an internal state sequence generated by a next-state function.

Theorem 1 *For any vectorial Boolean function $Z = f(X,Y)$ and any given Z, there exists a non-trivial algebraic equation over X iff $|f_x^{-1}(Z)| < 2^{n_x}$, and a polynomial $g \in \mathcal{P}_{n_x}^d$ defines a non-trivial algebraic equation over X iff the columns of $\mathbf{P}_{n_x}^d(f_x^{-1}(Z))$ corresponding to the monomials in g add up to zero. There exists a value of Z inducing a non-trivial algebraic equation over X of degree at most d if $n_z^* > n_{x,z}^* - n_x$ and*

$$\sum_{i=0}^{d} \binom{n_x}{i} > 2^{n_{x,z}^* - n_z^*} \tag{2}$$

and, in particular, if $n_z^* > n_y$ and

$$\sum_{i=0}^{d} \binom{n_x}{i} > 2^{n_x + n_y - n_z^*}. \tag{3}$$

If $n_z^* > n_y$ and n_z^* is an integer, then (3) holds if $d = \lceil n_x/2 \rceil$.

Proof A non-trivial algebraic equation over X conditioned on $f(X,Y) = Z$, for any given Z, is defined by a non-zero Boolean function $g(X)$, $X \in \{0,1\}^{n_x}$, such that $g(X) = 0$, $X \in f_x^{-1}(Z)$. The first claim then immediately follows from Lemma 1, by setting $n = n_x$ and $\mathcal{S} = f_x^{-1}(Z)$. By the same token, the second claim follows from Lemma 2.

The third claim is proven by applying Lemma 3. Namely, the total number of rows in all the matrices $\mathbf{P}_{n_x}^d(f_x^{-1}(Z))$ corresponding to $2^{n_z^*}$ different values of Z is exactly $2^{n_{x,z}^*}$. Therefore, among them there must exist a value of Z defining a matrix whose number of rows is at most $2^{n_{x,z}^* - n_z^*}$. By setting $\mathcal{S} = f_x^{-1}(Z)$, for such a Z, and by applying Lemma 3, we then obtain that there exists a non-trivial algebraic equation on $f_x^{-1}(Z)$ of degree at most d if (2) is satisfied, because of $2^{n_{x,z}^* - n_z^*} \geq |f_x^{-1}(Z)|$, which holds for such a Z. In addition, for (2) to hold, it is necessary that $n_{x,z}^* - n_z^* < n_x$.

The fourth claim is true due to $n_{x,z}^* < n_x + n_y$, whereas the last claim follows from the fact that if $n_z^* > n_y$ and n_z^* is an integer, then $n_z^* \geq n_y + 1$, so that $n_x + n_y - n_z^* \leq n_x - 1$. □

2.2 Unconditional Algebraic Equations

In the unconditional scenario, given a vectorial Boolean function $Z = f(X,Y)$, the objective is to find algebraic equations over X and Z. Variables in Y should thus be eliminated. This scenario is particularly interesting for the cryptanalysis of block and stream ciphers where f is an internal function whose output is unknown, e.g., corresponding to an intermediate round of a product block cipher or to an intermediate stage of an iterated construction of an output function in a stream cipher.

Theorem 2 *For any vectorial Boolean function $Z = f(X,Y)$, there exists a non-trivial algebraic equation over X and Z iff $n_{x,z}^* < n_x + n_z$, and a polynomial $g \in \mathcal{P}_{n_x+n_z}^d$ defines a non-trivial algebraic equation over X and Z iff the columns of $\mathbf{P}_{n_x+n_z}^d(f_{x,z})$ corresponding to the monomials in g add up to zero. There exists an induced non-trivial algebraic* *equation over X and Z of degree at most d if $n_z > n^*_{x,z} - n_x$ and*

$$\sum_{i=0}^{d} \binom{n_x + n_z}{i} > 2^{n^*_{x,z}} \tag{4}$$

and, in particular, if $n_z > n_y$ and

$$\sum_{i=0}^{d} \binom{n_x + n_z}{i} > 2^{n_x+n_y}. \tag{5}$$

Also, if $n_z > n_y$, then (5) holds if $d = \lceil (n_x + n_z)/2 \rceil$.

Proof Now, we set $n = n_x + n_z$ and $\mathcal{S} = f_{x,z}$, where $|f_{x,z}| = 2^{n^*_{x,z}}$. Then the first three claims directly follow from Lemmas 1-3, respectively. Note that for (4) to hold, it is necessary that $n^*_{x,z} < n_x + n_z$. The fourth claim is due to $n^*_{x,z} < n_x + n_y$, whereas the last claim is a consequence of $n_x + n_y \leq n_x + n_z - 1$. □

In a special case, by assuming that X is empty ($n_x = 0$), we get $f_{x,z} = f_z$ and $n^*_{x,z} = n^*_z$, and Theorem 2 is then about induced algebraic equations over the output variables, in Z.

2.3 Constrained Unconditional Algebraic Equations

In the constrained unconditional scenario, given a vectorial Boolean function $Z = f(X, Y)$, the objective is to find algebraic equations over X and Z having certain specific properties. As before, variables in Y should be eliminated. In particular, we look for algebraic equations having any degree in Z such that the monomials with the highest degrees in X do not depend on Z. This scenario is interesting for the so-called fast algebraic attacks on stream ciphers [5], as the high-degree monomials in X can be eliminated by using the linear complexity properties of the corresponding stream cipher.

More precisely, let $\mathcal{P}^{e,d}_{n_x,n_z}$ denote the set of all the multivariate polynomials in X and Z, $g(X, Z)$, of degree at most d in X and of any degree in Z such that the monomials whose degrees in X are larger than e do not depend on Z, where $1 \leq e \leq d$. Let $\mathbf{P}^{e,d}_{n_x,n_z}(f_{x,z})$ denote a matrix whose rows and columns are indexed by the vectors from $f_{x,z}$ and by the monomials from $\mathcal{P}^{e,d}_{n_x,n_z}$, respectively, and whose entries are defined by evaluating the monomials on these vectors.

Theorem 3 *For any vectorial Boolean function $Z = f(X, Y)$ and $1 \leq e \leq d$, a polynomial $g(X, Z) \in \mathcal{P}^{e,d}_{n_x,n_z}$ defines a non-trivial algebraic equation over X and Z iff the columns of $\mathbf{P}^{e,d}_{n_x,n_z}(f_{x,z})$ corresponding to the monomials in $g(X, Z)$ add up to zero. There exists an induced non-trivial algebraic equation over X and Z defined by a polynomial $g(X, Z) \in \mathcal{P}^{e,d}_{n_x,n_z}$ if $n_z > n^*_{x,z} - n_x$ and*

$$\sum_{i=0}^{d} \binom{n_x}{i} + (2^{n_z} - 1) \sum_{i=0}^{e} \binom{n_x}{i} > 2^{n^*_{x,z}} \tag{6}$$

and, in particular, if $n_z > n_y$ *and*

$$\sum_{i=0}^{d} \binom{n_x}{i} + (2^{n_z} - 1) \sum_{i=0}^{e} \binom{n_x}{i} > 2^{n_x + n_y}. \tag{7}$$

Furthermore, if $n_z^* = n_z$*, then there exists a value of* Z *for which such a polynomial* $g(X, Z)$ *is a non-zero polynomial in* X*.*

Proof As in the proof of Theorem 2, we set $n = n_x + n_z$ and $\mathcal{S} = f_{x,z}$, where $|f_{x,z}| = 2^{n^*_{x,z}}$. However, instead of considering the algebraic equations defined by $g(X, Z) \in \mathcal{P}^d_{n_x+n_z}$, we now consider the algebraic equations defined by $g(X, Z) \in \mathcal{P}^{e,d}_{n_x,n_z}$, where the number of monomials from $\mathcal{P}^{e,d}_{n_x,n_z}$ is equal to $\sum_{i=0}^{d} \binom{n_x}{i} + (2^{n_z} - 1) \sum_{i=0}^{e} \binom{n_x}{i}$.

Then the first two claims follow from analogs of Lemmas 2 and 3 pertaining to $\mathbf{P}^{e,d}_{n_x,n_z}(f_{x,z})$ and $\mathcal{P}^{e,d}_{n_x,n_z}$, respectively. As in Theorem 2, for (6) to hold, it is necessary that $n^*_{x,z} < n_x + n_z$, whereas the third claim is due to $n^*_{x,z} < n_x + n_y$. The last claim is proven by contraposition. Namely, if $n_z^* = n_z$ and for each $Z \in \{0,1\}^{n_z}$, the polynomial $g(X, Z)$ is equal to zero for every $X \in \{0,1\}^{n_x}$, then it follows that $g(X, Z)$ is the zero polynomial in X and Z, which contradicts the assumption. □

3 Algebraic Immunity Order

Consider a vectorial Boolean function $f : \{0,1\}^n \to \{0,1\}^m$, denoted as $Z = f(X)$, and let $m^* = \log |f_z|$, where f_z is the range of f. Let X' denote a generic subset of variables in X of size $n_{x'} = |X'|$ and let $X \setminus X'$ denote the subset of the remaining variables in X having size $n - n_{x'}$. We can then write $Z = f(X', X \setminus X')$ and consider algebraic equations involving X' that are induced by f.

Theorem 1 then implies that there are no algebraic equations over X' induced by any value of Z iff the set $\{f(X', X \setminus X') : X \setminus X' \in \{0,1\}^{n_x - n_{x'}}\}$ does not depend on X', that is, iff for each fixed value of X', the set of output values of $f(X', X \setminus X')$, for all different values of $X \setminus X'$, is the same. This justifies the following definition.

Definition 1 *The maximal* k *such that for every subset* X' *of size* $n_{x'} = |X'| = k$*, the set* $\{f(X', X \setminus X') : X \setminus X' \in \{0,1\}^{n_x - n_{x'}}\}$ *is the same for each value of* X' *is called the algebraic resiliancy order of* f*. In particular, if* $m^* = m \leq n$*, that is, if the range of* f *is maximal, then such* k *is called the algebraic immunity order.*

Let k denote the algebraic resiliancy order of f.

Proposition 1 *There exists a subset* X' *of size* $k + 1$ *and an output value* Z *such that there is a non-trivial algebraic equation over* X' *induced by* Z*.*

Proposition 2 $0 \leq k \leq n - m^*$ *and* k *cannot be smaller than the correlation resiliancy/immunity order of* f*. (Recall that the correlation resiliancy/immunity order of* f *is defined analogously* [13]*, with a difference that* $\{f(X', X \setminus X') : X \setminus X' \in \{0,1\}^{n_x - n_{x'}}\}$ *is regarded as a multiset, so that it is the distribution of output values that matters, not only the output values themselves.)*

Proposition 3 *If $m^* = 0$, i.e., if f is a constant function, then $k = n$. If f is a (non-constant) Boolean function effectively depending on all n input variables, then $m = m^* = 1$ and $k = n - 1$ holds iff f is affine, i.e., iff the algebraic degree of f is 1.*

Proof The first claim directly follows from the definition. As for the second claim, if we fix the values of any subset of $n - 1$ variables and vary the remaining variable, then the corresponding restriction of f attains both output values iff it is affine in this variable. Further, if for any fixed value of the chosen subset of $n - 1$ variables, the corresponding restriction of f is affine in the remaining variable, then f is affine in this variable. As a consequence, if $k = n - 1$, then f is affine, while the converse is trivially true, provided that f effectively depends on all n variables. □

So, if a Boolean function is not affine, that is, if the algebraic degree of f is at least 2, then $0 \leq k \leq n - 2$. Except in this particular case, there is no tradeoff between the algebraic resiliancy order and algebraic degree of f similar to that between correlation immunity order and algebraic degree, due to a less restrictive definition.

Definition 2 *For each $k + 1 \leq i \leq n$, we can find the minimal degree, d_i, of induced non-trivial algebraic equations over all subsets X' of size i and over all output values Z. The resulting sequence $(d_i)_{i=k+1}^n$ is non-increasing and is called the induced algebraic degree profile of f. The minimal value, d_n, is called the induced algebraic degree, $\tilde{d}$, of f.*

The name induced algebraic degree seems to be more appropriate than the name algebraic immunity introduced in [11].

From Theorem 1, we obtain the following property.

Proposition 4 *The induced algebraic degree is smaller than or equal to the minimal value of d satisfying*

$$\sum_{i=0}^{d} \binom{n}{i} > 2^{n-m^*}. \tag{8}$$

For non-constant Boolean functions, $m^* = 1$ and hence $\tilde{d} \leq \lceil n/2 \rceil$ [4], [11]. For invertible functions, $m^* = m$ and hence $\tilde{d} = 1$. This is natural as any known output value uniquely determines the input value, and that can be expressed by linear equations in the input variables. In general, an application of the well-known inequality

$$\frac{1}{\sqrt{2n}} 2^{nH(d/n)} \leq \frac{1}{\sqrt{8d(n-d)/n}} 2^{nH(d/n)} \leq \sum_{i=0}^{d} \binom{n}{i} \leq 2^{nH(d/n)}, \tag{9}$$

holding for $d \leq n/2$, where $H(p) = -p \log p - (1 - p) \log(1 - p)$ is the binary entropy function, results in the following numerical estimate of $\tilde{d}$, for $m^* \geq (1 + \log n)/2$,

$$\tilde{d} \leq \left\lfloor nH^{-1}\left(\frac{n - m^* + (1 + \log n)/2}{n}\right)\right\rfloor + 1. \tag{10}$$

For cryptographic functions, it is reasonable to require that the algebraic resiliancy order is close to maximum, $n - m^*$, and that the induced algebraic degree is also close to maximum, which is the minimal value of d satisfying (8).

4 Combiners with Memory

A binary combiner with k inputs, m outputs, and l bits of memory is a non-autonomous finite-state machine defined by $Y_{t+1} = g(X_t, Y_t)$ and $Z_t = h(X_t, Y_t)$, $t \geq 1$, where $g : \{0,1\}^k \times \{0,1\}^l \to \{0,1\}^l$ is a next-state vectorial Boolean function, $h : \{0,1\}^k \times \{0,1\}^l \to \{0,1\}^m$ is an output vectorial Boolean function, $Y_t = (y_{1,t}, \cdots, y_{l,t})$ is a state (memory) vector at time t, Y_1 is an initial state, $X_t = (x_{1,t}, \cdots, x_{k,t})$ is an input vector at time t and $Z_t = (z_{1,t}, \cdots, z_{m,t})$ is the output vector at time t. The input sequences are generated by a linear autonomous finite-state machine, typically, by k linear feedback shift registers (LFSR's).

Our objective is to find induced algebraic equations over the input sequences when the output sequences are assumed to be known, in the conditional scenario, and over the input and output sequences, in the unconditional scenario. In both scenarios, the state sequence has to be eliminated from the equations. This is achieved by studying the associated vectorial Boolean function $f_r : \{0,1\}^{kr} \times \{0,1\}^l \to \{0,1\}^{mr}$, for a generic $r \geq 1$, that defines a block of r consecutive outputs as a function of the corresponding block of r consecutive inputs and the preceding state, that is, $Z^r = f_r(X^r, Y)$. Namely, if $Z_t^r = (Z_t, \cdots, Z_{t+r-1})$ and $X_t^r = (X_t, \ldots, X_{t+r-1})$ denote blocks of r consecutive output and input vectors starting at time t, respectively, then it follows that $Z_t^r = f_r(X_t^r, Y_t)$, $t \geq 1$, where f_r is independent of t.

It is readily seen that one can directly apply theorems from Section 2 to the vectorial Boolean function f_r, by setting $X = X^r$, $Y = Y$, $Z = Z^r$, and $f = f_r$. Consequently, $n_x = kr$, $n_y = l$, and $n_z = mr$, while the value of n_z^*, denoted as $m_r = \log|\{Z^r : (\exists\, X^r, Y)\, f_r(X^r, Y) = Z^r\}|$, can depend on r in various ways. In particular, to satisfy $m_r = mr$ for any r, it is necessary that $k \geq m$. Induction over r immediately shows that to satisfy $m_r = mr$ for each r, it is sufficient that the output function h achieves all 2^m values for any fixed value of the state vector. This is a generalization of the condition from [1], which holds for $m = 1$. Note that for $f = f_r$ and linear algebraic equations, the set $f_x^{-1}(Z)$, the matrix $\mathbf{P}^1_{n_x}(f_x^{-1}(Z))$, and an application of the Gaussian elimination method are proposed in [8], when considering combiners with memory and single or multiple outputs.

For simplicity, we concentrate only on the sufficient conditions given in Theorems 1-3, but also bear in mind the necessary and sufficient conditions from these theorems, which show that the induced algebraic equations may also exist if these sufficient conditions are not satisfied (see Section 3).

Theorem 4 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$, there exists a value of Z^r inducing a non-trivial algebraic equation over X^r of degree at most d if $m_r > l$ and*

$$\sum_{i=0}^{d} \binom{kr}{i} > 2^{kr+l-m_r}. \tag{11}$$

If $k \geq m$, $m_r = mr$, and $r \geq \lceil (l+1)/m \rceil$, then (11) *holds if $d = \lceil kr/2 \rceil$.*

Theorem 4 is a generalization of the correponding theorem from [1], which holds for $m = 1$ and $m_r = r$. This theorem itself generalizes an earlier theorem from [4], which holds for $l = 0$. Note that the corresponding theorem from [6], holding for any l and $m \geq 1$, is not a proper generalization of the theorem from [1] as it holds in the unconditional scenario where the algebraic degree in Z^r can be arbitrary and as such does not guarantee the existence of a value of Z^r yielding a non-trivial algebraic equation only in X. In this regard, Theorem 4 in fact verifies in a mathematically precise way the conjecture from [6] that a non-trivial algebraic equation exists if "the output bits are fully independent and not related by some algebraic equation and if the output takes all possible 2^m values".

Theorem 5 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$, there exists an induced non-trivial algebraic equation over X^r and Z^r of degree at most d if $r \geq \lceil (l+1)/m \rceil$ and*

$$\sum_{i=0}^{d} \binom{(k+m)r}{i} > 2^{kr+l}. \tag{12}$$

In particular, if $r \geq \lceil (l+1)/m \rceil$, then (12) holds if $d = \lceil (k+m)r/2 \rceil$.

Theorem 6 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$ and for $1 \leq e \leq d$, there exists an induced non-trivial algebraic equation over X^r and Z^r defined by a polynomial $g(X^r, Z^r)$ of degree at most d in X^r and of any degree in Z^r, such that the monomials whose degrees in X^r are larger than e do not depend on Z^r, if $r \geq \lceil (l+1)/m \rceil$ and*

$$\sum_{i=0}^{d} \binom{kr}{i} + (2^{mr} - 1) \sum_{i=0}^{e} \binom{kr}{i} > 2^{kr+l}. \tag{13}$$

Furthermore, if $k \geq m$ and $m_r = mr$, then there exists a value of Z^r for which such a polynomial $g(X^r, Z^r)$ is a non-zero polynomial in X^r.

Theorem 6 is a generalization of the corresponding theorem from [5], which holds for $l = 0$, $m = 1$, and $r = 1$, that is, for a memoryless combiner with a Boolean output function. In this case, for (13) to hold, it is sufficient that $e + d \geq k$. Note that the last claim in Theorem 6 is new and is important as it establishes the existence of a non-trivial algebraic equation only in X.

5 Combiners with Finite Input Memory

A combiner is said to have finite input memory if the state vector is composed of a finite number of consecutive preceding bits in each of the input sequences, so that current output bits can be expressed as a function of the current input bits and a finite number of preceding bits in each of the input sequences. Examples include a well-known binary nonlinear filter generator, which can be regarded as a binary combiner with a finite input memory and with one input and one or more outputs and with and, more generally, a binary combiner composed of a number of LFSR's in which the outputs of a number of stages from each of the LFSR's are taken to the input of a vectorial Boolean function to produce the output.

In this case, the induced algebraic equations are allowed to involve the state bits too, so that non-trivial equations exist for any $r \geq 1$ and their degrees can be smaller. In view of (9) and (10), an estimate of the reduced degree in terms of the binary entropy function is also provided.

Theorem 7 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$ such that $m_r \geq 1$ and for any $r \geq 1$, there exists a value of Z^r inducing a non-trivial algebraic equation over X^r and Y of degree at most d if*

$$\sum_{i=0}^{d} \binom{kr+l}{i} > 2^{kr+l-m_r} \tag{14}$$

and, in particular, if $d = \lceil (kr+l)/2 \rceil$. If $k \geq m$ and $m_r = mr$, then (14) holds if

$$d > (kr+l)\, H^{-1}\left(\frac{(k-m)r + l + (1+\log(kr+l))/2}{kr+l}\right) \tag{15}$$

provided that $mr \geq (1+\log(kr+l))/2$.

Theorem 8 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$ and for any $r \geq 1$, there exists an induced non-trivial algebraic equation over X^r, Y, and Z^r of degree at most d if*

$$\sum_{i=0}^{d} \binom{(k+m)r+l}{i} > 2^{kr+l} \tag{16}$$

and, in particular, if $d = \lceil ((k+m)r+l)/2 \rceil$. Also, (16) holds if

$$d > ((k+m)r+l)\, H^{-1}\left(\frac{kr + l + (1+\log((k+m)r+l))/2}{(k+m)r+l}\right) \tag{17}$$

provided that $mr \geq (1+\log((k+m)r+l))/2$.

Theorem 9 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X^r, Y)$, $1 \leq e \leq d$, and for any $r \geq 1$, there exists an induced non-trivial algebraic equation over X^r, Y, and Z^r defined by a polynomial $g(X^r, Y, Z^r)$ of degree at most d in X^r and Y and of any degree in Z^r, such that the monomials whose degrees in X^r and Y are larger than e do not depend on Z^r, if*

$$\sum_{i=0}^{d} \binom{kr+l}{i} + (2^{mr}-1) \sum_{i=0}^{e} \binom{kr+l}{i} > 2^{kr+l}. \tag{18}$$

Furthermore, if $k \geq m$ and $m_r = mr$, then there exists a value of Z^r for which such a polynomial $g(X^r, Y, Z^r)$ is a non-zero polynomial in X^r and Y.

6 Divide-and-Conquer Algebraic Equations

If a binary combiner with or without memory consists of a number of different LFSR's, then the objective of algebraic attacks may be to reconstruct the initial states of a chosen subset of LFSR's by using the induced algebraic equations involving only the targeted LFSR sequences. This divide-and-conquer effect may thus significantly reduce the complexity of algebraic attacks. Accordingly, it is interesting to study the conditions under which such algebraic equations exist by applying theorems from Section 2 to essentially the same vectorial Boolean functions as in Sections 4 and 5, but with a difference that some input variables have to be eliminated from the induced algebraic equations.

As in Sections 4 and 5, we also concentrate only on the sufficient conditions given in Theorems 1-3, but bear in mind the necessary and sufficient conditions from these theorems, which show that the induced algebraic equations may also exist if these sufficient conditions are not satisfied. Theorems from Sections 6.1 and 6.2 can be regarded as generalizations of the correpsonding theorems from Sections 4 and 5, respectively. With respect to the sufficient conditions, it thus turns out that multiple outputs enable a divide-and-conquer effect which facilitates algebraic attacks. However, even if a combiner has a single binary output, a divide-and-conquer effect may be achievable. For example, it is shown in Section 3 that if a Boolean function of k input variables is not affine, then there must exist a subset of $k-1$ input variables allowing a non-trivial algebraic equation.

6.1 Combiners with Memory

Consider a binary combiner with k inputs, m outputs, and l bits of memory, with a modified objective to find induced algebraic equations involving only a subset of k' input sequences, $1 \leq k' \leq k$. We then have to study the same associated vectorial Boolean function $Z^r = f_r(X^r, Y)$, but with an appropriate partition of input variables. Namely, let X'^r denote the vector of input variables corresponding to the chosen subset of k' input sequences, where $|X'^r| = k'r$, and let $X^r \setminus X'^r$ denote the vector of the remaining input variables. We can then write $Z^r = f_r(X'^r, X^r \setminus X'^r, Y)$ and consider algebraic equations involving X'^r that are induced by f_r, where now both Y and $X^r \setminus X'^r$ have to be eliminated from the equations.

Theorems from Section 2 now have to be applied to the vectorial Boolean function f_r, by setting $X = X'^r$, $Y = (X^r \setminus X'^r, Y)$, $Z = Z^r$, and $f = f_r$. Consequently, $n_x = k'r$ and $n_y = (k-k')r + l$, whereas, as before, $n_z = mr$, and $n_z^* = m_r$.

Theorem 10 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, X^r \setminus X'^r, Y)$, where $|X'^r| = k'r$, there exists a value of Z^r inducing a non-trivial algebraic equation over X'^r of degree at most d if $m_r - (k-k')r > l$ and*

$$\sum_{i=0}^{d} \binom{k'r}{i} > 2^{kr+l-m_r}. \tag{19}$$

If $k \geq m > k-k'$, $m_r = mr$, and $r \geq \lceil (l+1)/(m-k+k') \rceil$, then (19) *holds if $d = \lceil k'r/2 \rceil$.*

Theorem 11 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, X^r \setminus X'^r, Y)$, where $|X'^r| = k'r$, there exists an induced non-trivial algebraic equation over X'^r and Z^r of degree at most d if $m > k - k'$, $r \geq \lceil (l+1)/(m-k+k') \rceil$, and*

$$\sum_{i=0}^{d} \binom{(k'+m)r}{i} > 2^{kr+l}. \tag{20}$$

In particular, if $m > k - k'$ and $r \geq \lceil (l+1)/(m-k+k') \rceil$, then (20) holds if $d = \lceil (k'+m)r/2 \rceil$.

Theorem 12 *For any binary combiner with k inputs, m outputs, and l bits of memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, X^r \setminus X'^r, Y)$, where $|X'^r| = k'r$, and for $1 \leq e \leq d$, there exists an induced non-trivial algebraic equation over X'^r and Z^r defined by a polynomial $g(X'^r, Z^r)$ of degree at most d in X'^r and of any degree in Z^r, such that the monomials whose degrees in X'^r are larger than e do not depend on Z^r, if $m > k - k'$, $r \geq \lceil (l+1)/(m-k+k') \rceil$, and*

$$\sum_{i=0}^{d} \binom{k'r}{i} + (2^{mr}-1) \sum_{i=0}^{e} \binom{k'r}{i} > 2^{kr+l}. \tag{21}$$

Furthermore, if $k \geq m > k - k'$ and $m_r = mr$, then there exists a value of Z^r for which such a polynomial $g(X'^r, Z^r)$ is a non-zero polynomial in X'^r.

6.2 Combiners with Finite Input Memory

Consider a binary combiner with k inputs, m outputs, and l bits of finite input memory. If the objective is to find induced algebraic equations involving only a subset of k' input sequences, $1 \leq k' \leq k$, then then the algebraic equations can also involve the l', $0 \leq l' \leq l$, state bits corresponding to these k' input sequences. Namely, we can then write $Z^r = f_r(X'^r, Y', X^r \setminus X'^r, Y \setminus Y')$, where $|X'^r| = k'r$, $|Y'| = l'$, and both $X^r \setminus X'^r$ and $Y \setminus Y'$ have to be eliminated from the equations.

Theorem 13 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, Y', X^r \setminus X'^r, Y \setminus Y')$, where $|X'^r| = k'r$ and $|Y'| = l'$, there exists a value of Z^r inducing a non-trivial algebraic equation over X'^r and Y' of degree at most d if $m_r - (k-k')r > l - l'$ and*

$$\sum_{i=0}^{d} \binom{k'r + l'}{i} > 2^{kr+l-m_r}. \tag{22}$$

If $k \geq m > k - k'$, $m_r = mr$, and $r \geq \lceil (l - l' + 1)/(m-k+k') \rceil$, then (22) holds if $d = \lceil (k'r + l')/2 \rceil$ as well as if

$$d > (k'r + l')\, H^{-1}\left(\frac{(k-m)r + l + (1 + \log(k'r + l'))/2}{k'r + l'} \right) \tag{23}$$

provided that $(m-k+k')r \geq l - l' + (1 + \log(k'r + l'))/2$.

Theorem 14 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, Y', X^r\backslash X'^r, Y\backslash Y')$, where $|X'^r| = k'r$ and $|Y'| = l'$, there exists an induced non-trivial algebraic equation over X'^r, Y', and Z^r of degree at most d if $m > k - k'$, $r \geq \lceil (l - l' + 1)/(m - k + k') \rceil$, and*

$$\sum_{i=0}^{d} \binom{(k' + m)r + l'}{i} > 2^{kr+l}. \tag{24}$$

If $m > k - k'$ and $r \geq \lceil (l - l' + 1)/(m - k + k') \rceil$, then (24) holds if $d = \lceil ((k' + m)r + l')/2 \rceil$ as well as if

$$d > ((k' + m)r + l')\, H^{-1} \left(\frac{kr + l + (1 + \log((k' + m)r + l'))/2}{(k' + m)r + l'} \right) \tag{25}$$

provided that $(m - k + k')r \geq l - l' + (1 + \log(k'r + l'))/2$.

Theorem 15 *For any binary combiner with k inputs, m outputs, and l bits of finite input memory having the associated vectorial Boolean function $Z^r = f_r(X'^r, Y', X^r\backslash X'^r, Y\backslash Y')$, where $|X'^r| = k'r$ and $|Y'| = l'$, and for $1 \leq e \leq d$, there exists an induced non-trivial algebraic equation over X'^r, Y', and Z^r defined by a polynomial $g(X'^r, Y', Z^r)$ of degree at most d in X'^r and Y' and of any degree in Z^r, such that the monomials whose degrees in X'^r and Y' are larger than e do not depend on Z^r, if $m > k - k'$, $r \geq \lceil (l - l' + 1)/(m - k + k') \rceil$, and*

$$\sum_{i=0}^{d} \binom{k'r + l'}{i} + (2^{mr} - 1) \sum_{i=0}^{e} \binom{k'r + l'}{i} > 2^{kr+l}. \tag{26}$$

Furthermore, if $k \geq m > k - k'$ and $m_r = mr$, then there exists a value of Z^r for which such a polynomial $g(X'^r, Y', Z^r)$ is a non-zero polynomial in X'^r and Y'.

7 Conclusions

Some general concepts and results on vectorial Boolean functions and induced algebraic equations are presented. The conditional scenario considered is more useful for stream ciphers, whereas the unconditional scenario may be useful for product block ciphers and for stream ciphers whose output function depends on a large number of state variables, and is obtained by an iterated construction. The constrained unconditional scenario is useful for the so-called fast algebraic attacks on stream ciphers. The notion of correlation resiliancy/immunity order is extended to that of algebraic resiliancy/immunity order and some basic properties are derived.

The new contributions on combiners with memory include differentiating among the three scenarios, showing the importance of the range of the underlying vectorial Boolean function, proving that a finite input memory (e.g., in a nonlinear filter generator) can considerably reduce the induced algebraic degree, and pointing out a possible divide-and-conquer effect, especially in the case of multiple binary outputs, which can significantly reduce the complexity of algebraic attacks. As the algorithms for finding induced algebraic equations of low degree are exponential in the number of input variables, it is interesting to look for more efficient algorithms, not only in the case of algebraic equations holding with certainty, but also in a more general case where the equations are allowed to hold with a high probability rather than with certainty.

References

[1] F. Armknecht and M. Krause, "Algebraic attacks on combiners with memory," Advances in Cryptology - Crypto 2003, *Lecture Notes in Computer Science*, vol. 2729, pp. 162-176, 2003.

[2] N. Courtois, "Higher order correlation attacks, XL algorithm and cryptanalysis of Toyocrypt," ICISC 2002, *Lecture Notes in Computer Science*, vol. 2587, pp. 182-199, 2002.

[3] N. Courtois and J. Pieprzyk, "Cryptanalysis of block ciphers with overdefined systems of equations," Advances in Cryptology - Asiacrypt 2002, *Lecture Notes in Computer Science*, vol. 2501, pp. 267-287, 2002.

[4] N. Courtois and W. Meier, "Algebraic attacks on stream ciphers with linear feedback," Advances in Cryptology - Eurocrypt 2003, *Lecture Notes in Computer Science*, vol. 2656, pp. 345-359, 2003.

[5] N. Courtois, "Fast algebraic attacks on stream ciphers with linear feedback," Advances in Cryptology - Crypto 2003, *Lecture Notes in Computer Science*, vol. 2729, pp. 176-194, 2003.

[6] N. Courtois, "Algebraic attacks on combiners with memory and several outputs," preprint, June 2004.

[7] J. Dj. Golić, "Correlation properties of a general binary combiner with memory," *Journal of Cryptology*, vol. 9(2), pp. 111-126, 1996.

[8] Jovan Dj. Golić, "Conditional correlation attack on combiners with memory," *Electronics Letters*, vol. 32(24), pp. 2193-2195, 1996.

[9] P. Hawkes and G. Rose, "Rewriting variables: the complexity of fast algebraic attacks on stream ciphers," Advances in Cryptology - Crypto 2004, *Lecture Notes in Computer Science*, vol. 3152, pp. 390-406, 2004.

[10] W. Meier and O. Staffelbach, "Correlation properties of combiners with memory in stream ciphers," *Journal of Cryptology*, vol. 5(1), pp. 67-86, 1992.

[11] W. Meier, E. Pasalic, and C. Carlet, "Algebraic attacks and decomposition of Boolean functions," Advances in Cryptology - Eurocrypt 2004, *Lecture Notes in Computer Science*, vol. 3027, pp. 474-491, 2004.

[12] A. Shamir, J. Patarin, N. Courtois, and A. Klimov, "Efficient algorithms for solving overdefined systems of multivariate polynomial equations," Advances in Cryptology - Eurocrypt 2000, *Lecture Notes in Computer Science*, vol. 1807, pp. 392-407, 2000.

[13] T. Siegenthaler, "Correlation immunity of nonlinear combining functions for cryptographic applications," *IEEE Trans. Inform. Theory*, vol. 30, pp. 776-780, 1984.