On a weakness of the Tu-Deng function and its repair

Claude Carlet *

Abstract

We observe that the function introduced by Z. Tu and Y. Deng in the ePrint Archive paper 2009/272 is weak against fast algebraic attacks. We propose an alternative function sharing all the properties of the Tu-Deng function but having not this weakness.

Keywords: Algebraic attack, Boolean function, Stream cipher

1 Introduction

For being usable as a filter function in the pseudo-random generator of a stream cipher, a Boolean function $f : F_2^n \mapsto F_2$ must be balanced (that is, have a uniformly distributed output, to avoid distinguishing attacks), have high nonlinearity (that is, lie at a Hamming distance from affine Boolean functions close to the maximum $2^{n-1} - 2^{n/2-1}$, to allow resistance to fast correlation attacks [10]), have high algebraic degree (i.e. have a polynomial representation in $F_2[x_1, \cdots, x_n]/(x_1 + x_1^2, \cdots, x_n + x_n^2)$ with degree close to the maximum n, to avoid Berlekamp-Massey attack [9] and Rønjom-Helleseth attack [12]), have high algebraic immunity (close to the maximum $\lceil n/2 \rceil$ to avoid the standard algebraic attack [5, 11]) and behave well against fast algebraic attacks [4]. If one of these requirements is not fulfilled by a function, then the function may represent a theoretical interest but is not suitable for applications in stream ciphers.

In [3], an infinite class of balanced functions in any number n of variables, with optimal algebraic immunity has been introduced. The algebraic degree of the functions in this class is optimal and its nonlinearity computed for small values of n is very good; a lower bound on the nonlinearity could be proved and shows that the nonlinearity is much better than for all the previously introduced functions with high algebraic immunity but this lower bound is not enough for asserting that the function allows resisting the fast correlation attack; the gap between the computed values and the provable lower bound for this kind of

*LAGA, University of Paris 8, 2 rue de la liberté, 93526 Saint-Denis, Cedex France; Email: claude.carlet@inria.fr.

functions has been also observed by researchers working on sequences (see e.g. [6]).

In [13] is introduced another infinite class of balanced functions, in any even number of variables, with optimal algebraic immunity, which is related to the class of [3]. The nonlinearity of the functions in this class is provably good, as well as the algebraic degree. The optimal algebraic immunity of the function is proved in [13] up to a conjecture which has been proved til $n/2 = 29$, which is quite sufficient for cryptographic purposes. So this class seemed promising.

However we point out in the present note that the functions of this class are weak against fast algebraic attacks. We propose another balanced function, having not this weakness, and having optimal algebraic immunity and algebraic degree and good nonlinearity.

2 Preliminaries

Let $\mathbb{F}_2^n$ be the n-dimensional vector space over $\mathbb{F}_2$, and B_n the set of n-variable (Boolean) functions from $\mathbb{F}_2^n$ to $\mathbb{F}_2$. The basic representation of a Boolean function $f(x_1, \cdots, x_n)$ is by the output column of its truth table, i.e., a binary string of length 2^n,

$$[f(0,0,\cdots,0), f(1,0,\cdots,0), f(0,1,\cdots,0), f(1,1,\cdots,0), \cdots, f(1,1,\cdots,1)].$$

The *Hamming weight* $\mathrm{wt}(f)$ of a Boolean function $f \in B_n$ is the weight of this string, that is, the size of the support $\mathrm{Supp}(f) = \{x \in \mathbb{F}_2^n \,|\, f(x) = 1\}$ of the function. The *Hamming distance* $\mathrm{d_H}(f, g)$ between two Boolean functions f and g is the Hamming weight of their difference $f + g$ (by abuse of notation, we use $+$ to denote the addition on $\mathbb{F}_2$, i.e., the XOR). We say that a Boolean function f is *balanced* if its truth table contains an equal number of 1's and 0's, that is, if its Hamming weight equals 2^{n-1}.

Any Boolean function has a unique representation as a multivariate polynomial over $\mathbb{F}_2$, called the *algebraic normal form* (ANF), of the special form:

$$f(x_1, \cdots, x_n) = \sum_{I \subseteq \{1,2,\cdots,n\}} a_I \prod_{i \in I} x_i.$$

The *algebraic degree*, $\deg(f)$, is the global degree of this polynomial, that is, the number of variables in the highest order term with non zero coefficient. A Boolean function is affine if it has degree at most 1. The set of all affine functions is denoted by A_n.

We shall need another representation of Boolean functions, by univariate polynomials over the field $\mathbb{F}_{2^n}$. We identify the field $\mathbb{F}_{2^n}$ and the vector space $\mathbb{F}_2^n$: this field being an n-dimensional $\mathbb{F}_2$-vector space, we can choose a basis $(\beta_1, \cdots, \beta_n)$ and identify every element $x = \sum_{i=1}^n x_i \beta_i \in \mathbb{F}_{2^n}$ with the n-tuple of its coordinates $(x_1, \cdots, x_n) \in \mathbb{F}_2^n$. Every function $f : \mathbb{F}_{2^n} \to \mathbb{F}_{2^n}$ (and in particular every Boolean function $f : \mathbb{F}_{2^n} \to \mathbb{F}_2$) can then be uniquely represented as a polynomial $\sum_{j=0}^{2^n-1} a_j x^j$ where $a_j \in \mathbb{F}_{2^n}$. Indeed, the mapping which

maps every such polynomial to the corresponding function from $\mathbb{F}_{2^n}$ to itself is $\mathbb{F}_{2^n}$-linear, injective (since a non-zero polynomial of degree at most 2^n-1 over a field cannot have more than 2^n-1 zeroes in this field) and therefore surjective since the $\mathbb{F}_{2^n}$-vector spaces of these polynomials and of the functions from $\mathbb{F}_{2^n}$ to itself have the same dimension 2^n. The function is Boolean if and only if the functions $f(x)$ and $(f(x))^2$ are represented by the same polynomial, that is, if $a_0, a_{2^n-1} \in \mathbb{F}_2$ and, for every $i = 1, \cdots, 2^n-2$, we have $a_{2j} = (a_j)^2$, where $2j$ is taken mod 2^n-1. Then the algebraic degree of the function equals the maximum *2-weight* $w_2(j)$ of j such that $a_j \neq 0$, where the 2-weight of j equals the number of 1's in its binary expansion. In this representation, the elements of A_n are all the functions $tr(ax)$, $a \in \mathbb{F}_{2^n}$, where tr is the trace function: $tr(x) = x + x^2 + x^{2^2} + \cdots + x^{2^{n-1}}$.

Any Boolean function should have high algebraic degree to allow the cryptosystem resisting the Berlekamp-Massey attack [9].

Boolean functions used in cryptographic systems must have high nonlinearity to withstand fast correlation attacks (see e.g. [10]). The *nonlinearity* of an n-variable function f is its distance to the set of all n-variable affine functions, i.e.,

$$nl(f) = \min_{g \in A_n} (\mathrm{d_H}(f, g)).$$

This parameter can be expressed by means of the Walsh transform. Let $x = (x_1, \cdots, x_n)$ and $\lambda = (\lambda_1, \cdots, \lambda_n)$ both belong to $\mathbb{F}_2^n$ and $\lambda \cdot x$ be the usual inner product in $\mathbb{F}_2^n$: $\lambda \cdot x = \lambda_1 x_1 + \cdots + \lambda_n x_n \in \mathbb{F}_2$, or any other inner product in $\mathbb{F}_2^n$. Let $f(x)$ be a Boolean function in n variables. The *Walsh transform* (depending on the choice of the inner product) of $f(x)$ is the integer valued function over $\mathbb{F}_2^n$ defined as

$$W_f(\lambda) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x)+\lambda \cdot x}.$$

If we identify the vector space $\mathbb{F}_2^n$ with the field $\mathbb{F}_{2^n}$, then we can take for inner product: $\lambda \cdot x = tr(\lambda x)$.
A Boolean function f is balanced if and only if $W_f(0) = 0$. The nonlinearity of f can also be given by

$$nl(f) = 2^{n-1} - \frac{1}{2} \max_{\lambda \in \mathbb{F}_2^n} |W_f(\lambda)|.$$

For every n-variable function f we have $nl(f) \leq 2^{n-1} - 2^{n/2-1}$.

Algebraic attacks have been introduced recently (see [5]). They recover the secret key, or at least the initialization of the cipher, by solving a system of multivariate algebraic equations. In practice, for cryptosystems which are robust against the usual attacks, this system is too complex to be solved (its equations being highly nonlinear). In the case of stream ciphers, we can get a very overdefined system (i.e. a system with a number of linearly independent

equations much greater than the number of unknowns). In the combiner or the filter model, with a linear part of size N and with an n-variable Boolean function f as combining or filtering function, there exists a linear permutation $L : \mathbb{F}_2^N \mapsto \mathbb{F}_2^N$ and a linear mapping $L' : \mathbb{F}_2^N \mapsto \mathbb{F}_2^n$ such that, denoting by $u_1, \cdots, u_N$ the initialisation and by $(s_i)_{i\geq 0}$ the pseudo-random sequence output by the generator, we have, for every $i \geq 0$:

$$s_i = f(L' \circ L^i(u_1, \cdots, u_N)).$$

The number of equations can then be much larger than the number of unknowns. This makes less complex the resolution of the system by using Groebner basis, and even allows linearizing the system (i.e. obtaining a system of linear equations by replacing every monomial of degree greater than 1 by a new unknown); the resulting linear system has however too many unkwnowns and cannot be solved. Courtois and Meier have had a simple but very efficient idea. Assume that there exist functions $g \neq 0$ and h of low algebraic degrees (say, of degrees at most d) such that $f * g = h$. We have then, for every $i \geq 0$:

$$s_i \, g(L' \circ L^i(u_1, \cdots, u_N)) = h(L' \circ L^i(u_1, \cdots, u_N)).$$

This equation in $u_1, \cdots, u_N$ has degree at most d, since L and L' are linear, and the system of equations obtained after linearization can then be solved by Gaussian elimination. Low algebraic degree relations have been shown to exist for several well known constructions of stream ciphers, which were immune to all previously known attacks.
It has been shown [5, 11] that the existence of such relations is equivalent to that of non-zero functions g of low algebraic degrees such that $f * g = 0$ or $(f + 1) * g = 0$. This led to the following definition.

Definition 1 *For $f \in B_n$, we define $AN(f) = \{g \in B_n \,|\, f * g = 0\}$. Any function $g \in AN(f)$ is called an annihilator of f. The algebraic immunity (AI) of f is the minimum algebraic degree of all the nonzero annihilators of f and of all those of $f + 1$. We denote it by $AI(f)$.*

Note that $AI(f) \leq \deg(f)$, since $f*(1+f) = 0$. Note also that the algebraic immunity, as well as the nonlinearity and the algebraic degree, is affine invariant (i.e. is invariant under composition by an affine automorphism). As shown in [5], we have $AI(f) \leq \lceil \frac{n}{2} \rceil$.

The complexity of the standard algebraic attack on the combiner model or the filter model using a nonlinear function f equals roughly $O(D^3)$ in time and $O(D)$ in data, where $D = \sum_{i=0}^{AI(f)} \binom{N}{i}$, where N is the size of the linear part of the pseudo-random generator.

If a function has optimal algebraic immunity $\lceil \frac{n}{2} \rceil$ with n odd, then it is balanced. M. Lobanov has obtained in [8] the following tight lower bound:

$$nl(f) \geq 2 \sum_{i=0}^{AI(f)-2} \binom{n-1}{i}.$$

Note that this bound does not assure that the nonlinearity is high enough.

A high algebraic immunity is a necessary but not sufficient condition for robustness against all kinds of algebraic attacks. Indeed, if one can find g of low algebraic degree and $h \neq 0$ of reasonable algebraic degree such that $f * g = h$, then a fast algebraic attack (in brief FAA) is feasible, see [4], which has a pre-computation step of complexity $\mathrm{O}(Dlog^2D)$ and an on-line complexity of $\mathrm{O}(E^3 + EDlogD)$ where $E = \sum_{i=0}^{deg(g)} \binom{N}{i}$ (note that fast algebraic attacks need more data than standard ones). It has been shown in [4] that when $e + d \geq n$, there must exist g of algebraic degree at most e and h of algebraic degree at most d such that $f * g = h$. Hence, an n-variable function f can be considered as optimal with respect to fast algebraic attacks if there do not exist two functions $g \neq 0$ and h such that $f * g = h$ and $\deg(g) + \deg(h) < n$ with $\deg(g) < n/2$. Note that if we have $f * g = h$ with $g \neq 0$ and $deg(g) < AI(f)$ then we cannot have $h = 0$ (since g would then be a nonzero annihilator of f, a contradiction) and multiplying both terms of the equality by f we get $f * h = h$ and then h is a nonzero annihilator of $f+1$. Hence $deg(h) \geq AI(f)$. A function having algebraic immunity $\lceil n/2 \rceil$ is then the worst possible with respect to FAA if there exists a function g such that $deg(g) = 1$, and a function h such that $deg(h) = \lceil n/2 \rceil$ and $f * g = h$. The case which comes immediately after this worst case is when there exists g such that $deg(g) = 1$ and h of algebraic degree $deg(h) = \lceil n/2 \rceil + 1$ such that $f * g = h$.

The pseudo-random generator must also resist algebraic attacks on the augmented function [7] but this condition is more on the linear update function than on the filter function.

A new version of algebraic attack has been found recently by S. Rønjom and T. Helleseth [12] and is very efficient. Its time complexity is roughly $O(\mathcal{D}')$, where $\mathcal{D}' = \sum_{i=0}^{\deg(f)} \binom{N}{i}$, where N is the size of the linear part of the pseudo-random generator. But it needs much more data than standard algebraic attacks: $O(\mathcal{D}')$ also! When f has algebraic degree close to n and algebraic immunity close to $\frac{n}{2}$, this is the square of what is needed by standard algebraic attacks. However, this attack obliges the designer to choose a function with very high algebraic degree.

The functions used in the combiner model must be additionally highly resilient (that is, balanced and correlation immune of a high order; see definition *e.g.* in [2]) to withstand correlation attacks. It seems quite difficult to achieve all of the necessary criteria including this one, and for this reason, the filter generator seems more appropriate.

The balanced function g having optimal algebraic immunity introduced in [3] can be defined for every n as the Boolean function on $\mathbb{F}_{2^n}$ whose support equals $\{\alpha^i;\ i = s, \cdots, s + 2^{n-1} - 1\}$, where s is an integer and α is a primitive element of $\mathbb{F}_{2^n}$. Function g has algebraic degree $n - 1$ (which is optimal for a

balanced function) and its nolinearity satisfies:

$$nl(g) \geq 2^{n-1} + \frac{2^{n/2+1}}{\pi} \ln\left(\frac{\pi}{4(2^n-1)}\right) - 1 \approx 2^{n-1} - \frac{2\ln 2}{\pi}\, n\, 2^{n/2}.$$

The actual values of $nl(g)$ computed for small values of n happen to be much better than what gives this bound, but showing a bound proving that the nonlinearity of g is good is an open problem (also in sequence theory, see [6]). Concerning the behavior of the function against FAA, computer investigations made using [1, Algorithm 2] suggest the following properties of this class of functions:

- No nonzero function g' of algebraic degree at most e and no function h of algebraic degree at most d exist such that $g*g' = h$, when $(e,d) = (1, n-2)$ for n odd and $(e,d) = (1, n-3)$ for n even. This has been checked for $n \leq 12$ and we conjecture it for every n.
- For $e > 1$, pairs (g', h) of algebraic degrees (e,d) such that $e + d < n - 1$ were never observed. Precisely, the non-existence of such pairs could be checked exhaustively for $n \leq 9$ and $e < n/2$, for $n = 10$ and $e \leq 3$ and for $n = 11$ and $e \leq 2$. This suggests that this class of functions, even if not always optimal against fast algebraic attacks, has a very good behavior.

The instance with $n = 9$ turns out to be optimal. The problem of proving the good behavior of g against FAA for every n is open.
The complexity of computing $g(x)$ is same as for the discrete log, but n is small and the Pohlig-Hellman method can be used:

$$2^{18} - 1 = 27 * 73 * 133\ ; \quad 2^{20} - 1 = 41 * 93 * 275.$$

This allows computing 1 bit per cycle, using 40,000 transistors.

The Tu-Deng infinite class [13] of balanced functions with optimal algebraic immunity (if a conjecture can be proven) and provably good nonlinearity is the $2n$-variable function over $F_{2^n}^2$ defined as follows:

$$f(x,y) = \begin{cases} g\left(\frac{x}{y}\right) \text{ if } x \neq 0 \\ 1 \text{ if } x = 0 \text{ and } x \in \{\alpha^i;\ i = 2^{n-1} - 1, \cdots, 2^n - 2\} \\ 0 \text{ otherwise.} \end{cases}$$

It has optimal algebraic immunity as well and is balanced. Its algebraic degree equals $2n-1$ and its nonlinearity satisfies $nl(f) \geq 2^{2n-1} - 2^{n-1} - n\, 2^{n/2} \ln 2 - 1$.

3 The weakness of the Tu-Deng function against FAA

Let us denote by $b(x,y)$ the function $g(x/y)$ (with the convention that $x/0 = 0$ for every x). This is the function considered in Construction 1 of [13]. It is

bent and belongs to the well-known PSap Dillon's class of hyperbent functions (see e.g. [2]); it has algebraic degree exactly n (which is the highest possible degree for a bent function). We can see that function f differs from b only when $x = 0$. Hence for every linear Boolean function ℓ over F_{2^n}, the function $\ell(x)f(x,y)$ equals $\ell(x)b(x,y)$ and has therefore algebraic degree at most $n+1$. As recalled in Section 2, this is almost the worst case for the resistance to FAA of a $2n$-variable function of algebraic immunity n.

4 Repairing the Tu-Deng function

The observation in Section 3 shows that, to have a chance of allowing a good resistance to FAA, the function must differ from function b on any affine hyperplane of $F_2^{2n} \sim F_{2^n}^2$. In other words, the new function f, that we shall denote by f' for clarity, must be defined as

$$f'(x,y) = b(x,y) + 1_E(x,y)$$

where 1_E is the indicator of a set E spanning the whole F_2-vectorspace $F_{2^n}^2$ and such that, for every vector e, the set $e+E$ spans $F_{2^n}^2$ as well (this condition is indeed necessary and sufficient for g differing from b on any affine hyperplane since g equals b on the complement H^c of a linear hyperplane H if and only if E is included in H and g equals b on H if and only if $e+E$ is included in H, where $H^c = e+H$). We shall call (P) this property of E.
Let E be a set of the form

$$E = \{(0, u_\emptyset)\} \cup \left\{(\alpha^i u_i, u_i);\ i \notin \{s, \cdots, s+2^{n-1}-1\}\right\}$$

(α being the primitive element of F_{2^n} used for the definition of g and f) where for every i, the vector u_i is any nonzero element of F_{2^n}. We assume that E has the property (P). Since b has Hamming weight $2^{2n-1} - 2^{n-1}$ and E is disjoint from its support and has size 2^{n-1}, the function $f = b + 1_E$ is balanced. Since the support of f' includes the support of b and b has algebraic immunity n (see [13]), then f' has no nonzero annihilator of algebraic degree strictly less than n. Let $h(x,y) = \sum_{i=0}^{2^n-2}\sum_{j=0}^{2^n-2} a_{i,j}\ x^i y^j$ be a non-zero annihilator of $f'+1$ with $\max\{w_2(i)+w_2(j);\ a_{i,j} \neq 0\} \leq n-1$. We have $h(0,y) = 0$ and $h(\gamma\, y, y) = 0$ for every $y \in F_{2^n}^*$ and every $\gamma \notin supp(g) = \{\alpha^i;\ i = s, \cdots, s+2^{n-1}-1\}$, except maybe for $y = u_\emptyset$, resp. $y = u_i$.
For every $y \in F_{2^n}^*$, $h(\gamma\, y, y)$ equals:

$$\sum_{i=0}^{2^n-2}\sum_{j=0}^{2^n-2} a_{i,j}\ \gamma^i y^{i+j} = \sum_{t=0}^{2^n-2}\left(\sum_{i=0}^{2^n-2} a_{i,t-i}\ \gamma^i\right) y^t,$$

where $t-i$ is taken modulo 2^n-1. Since $h(\gamma y, y)$ is null for every $y \neq u_\emptyset$, resp. $y \neq u_i$, and has therefore at least 2^n-1 zeros and since $t \leq 2^n-2$ we have then $\sum_{i=0}^{2^n-2} a_{i,t-i}\ \gamma^i = 0$ for every t and the proof ends as in [13]. The

nonlinearity of f' satisfies the inequality $nl(f') \geq 2^{2n-1} - 2^n$ and maybe a better one for clever choices of E. We are currently making investigations to check that f' can have an optimal algebraic degree $2n-1$ and behave well against FAA.

Acknowledgement

We thank Lilya Budaghyan and Simon Fischer for their great help.

References

[1] Armknecht, F., Carlet, C., Gaborit, P., Künzli, S., Meier,W. and Ruatta, O.: Efficient computation of algebraic immunity for algebraic and fast algebraic attacks. Advances in Cryptology, EUROCRYPT 2006, Lecture Notes in Computer Science 4004 , pp. 147-164, 2006.

[2] Carlet, C. Boolean Functions for Cryptography and Error Correcting Codes. Chapter of the monography *Boolean Methods and Models*, Y. Crama and P. Hammer eds, Cambridge University Press, to appear. Preliminary version available at http://www-rocq.inria.fr/codes/Claude.Carlet/pubs.html

[3] Carlet, C. and Feng, K. An infinite class of balanced functions with optimum algebraic immunity, good immunity to fast algebraic attacks and good nonlinearity. *Proceedings of ASIACRYPT 2008, Lecture Notes in Computer Science* 5350, pp. 425-440, 2008.

[4] Courtois, N.: Fast algebraic attacks on stream ciphers with linear feedback, Advances in Cryptology–CRYPTO 2003, Lecture Notes in Computer Science, vol. 2729 pp. 176-194. Springer, Verlag (2003).

[5] Courtois, N., Meier, W.: Algebraic attacks on stream ciphers with linear feedback, Advances in Cryptology–Eurocrypt 2003, Lecture Notes in Computer Science, vol. 2656 pp. 345-359. Springer, Verlag (2003).

[6] Dmitriev, D. and Jedwab, J.. Bounds on the growth rate of the peak sidelobelevel of binary sequences. Advances in Mathematics of Communications, Volume 1, No. 4, 461475, 2007.

[7] Fischer, S. and Meier, W.: Algebraic Immunity of S-boxes and Augmented Functions. Proceedings of Fast Software Encryption 2007. Lecture Notes in Comput. Sci. 4593, pp. 366-381.

[8] Lobanov, M.: Tight bound between nonlinearity and algebraic immunity. Paper 2005/441 in http://eprint.iacr.org/

[9] Massey, J. L. Shift-register analysis and BCH decoding. *IEEE Transactions on Information Theory*, vol. 15, pp. 122-127, 1969.

[10] Meier, W. and Staffelbach, O. Fast correlation attacks on stream ciphers. *Advances in Cryptology, EUROCRYPT'88, Lecture Notes in Computer Science* 330, pp. 301-314, 1988.

[11] Meier, W., Pasalic, E., Carlet, C.: Algebraic attacks and decomposition of Boolean functions, Advances in Cryptology–EUROCRYPT 2004, Lecture Notes in Computer Science, vol. 3027 pp. 474-491. Springer, Verlag (2004).

[12] Rønjom, S., Helleseth, T.: A new attack on the filter generator. IEEE Trans. Inform. Theory 53(5) 1752-1758 (2007).

[13] Tu, Z. and Deng, Y. A Conjecture on Binary String and Its Applications on Constructing Boolean Functions of Optimal Algebraic Immunity. IACR ePrint Archive 2009/272.