A construction of Boolean functions with good cryptographic properties

Jong H. Chung, Pantelimon Stanica, Chik-How Tan, and Qichun Wang

1 Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943–5216, USA
\{jhchung, pstanica\}@nps.edu

2 Temasek Laboratories, National University of Singapore, 117411, Singapore
\{tslwq, tsltch\}@nus.edu.sg

Abstract.
The two important qualities of a cipher is security and speed. Frequently, to satisfy the security of a Boolean function primitive, speed may be traded-off. In this paper we present a general construction that addresses both qualities. The idea of our construction is to manipulate a cryptographically strong base function and one of its affine equivalent functions, using concatenation and negation. We achieve security from the inherent qualities of the base function, which are preserved (or increased) and obtain speed by the simple Boolean operations. We present two applications of the construction to demonstrate the flexibility and efficiency of the construction.

Keywords: Cryptographic Boolean functions, avalanche characteristics, resiliency, algebraic immunity, nonlinearity, hidden weighted bit function.

MSC 2010: 94C10, 94A60, 11T71

1 Introduction

A stream cipher typically employs at least one Linear Feedback Shift Register (LFSR) to generate a secret key stream. Due to the linear nature of LFSR, the raw bits from LFSRs can not be used to encrypt a message, since the encrypted message is vulnerable to linear cryptanalysis and other forms of attacks. Therefore, the stream cipher filters an LFSR or combines multiple LFSRs to generate stronger secret key stream bits, via a Boolean function as a nonlinear filter or combiner, respectively. Two key factors in designing cryptographic Boolean functions are security and speed. We achieve security by having good measures in as many cryptographic properties as possible for the Boolean function in a cipher, such as balancedness to resist statistical attacks, high nonlinearity to address linear cryptanalysis on block ciphers (or correlation attacks on stream ciphers), high algebraic degree against algebraic attacks (although, this is not sufficient), high correlation immunity and resiliency to deal with correlation attacks, and...
high algebraic immunity to resist (fast) algebraic attacks. Speed is another important aspect in a cipher, since we desire fast encryption and decryption. For example, the Carlet-Feng function has good cryptographic properties, but it is not simple to generate, which may affect certain ciphers to underperform. Here we present some ideas to construct good cryptographic Boolean functions using a cryptographically strong base function, and three simple Boolean operations, namely concatenation, affine transformation, and negation. One of the significant benefits from this construction is the flexibility to choose an appropriate base function with known cryptographic properties (and we propose several such). This means we can customize our function, focusing on certain cryptographic properties. The other benefit is that our function is easy to construct due to the three previously mentioned basic operations.

The paper is organized as follows. In Section 2, the necessary background is established. We then present a construction of functions based on concatenation, affine transformation, and negation in Section 3. In Section 4, we give applications for our construction.

2 Preliminaries

Let $F_2^n$ be the $n$-dimensional vector space over the binary field $F_2$, and $B_n$ the set of all $n$-variable Boolean functions defined on $F_2^n$.

Any Boolean function $f \in B_n$ can be uniquely represented as a multivariate polynomial in $F_2[x_1, \ldots, x_n]$, called the algebraic normal form (ANF),

$$f(x_1, \ldots, x_n) = \sum_{K \subseteq \{1,2,\ldots,n\}} a_K \prod_{k \in K} x_k.$$  

The algebraic degree of $f$, denoted by $\deg(f)$, is the number of variables in the highest order term with nonzero coefficients.

A Boolean function is affine if there exists no term of degree strictly greater than 1 in the ANF and the set of all affine functions is denoted by $A_n$. Let $f \in B_n$ and $E$ be any flat (that is, a coset of a vector subspace). If the restriction $f|_E$ of $f$ to $E$ is constant (respectively affine), then $E$ is called a constant (respectively affine) flat for $f$.

Let $1_f = \{x \in F_2^n | f(x) = 1\}$, $0_f = \{x \in F_2^n | f(x) = 0\}$, be the support of a Boolean function $f$, respectively, its complement. The cardinality of $1_f$ is called the Hamming weight of $f$, and will be denoted by $\text{wt}(f)$. The Hamming distance between two functions $f$ and $g$ is the Hamming weight of $f + g$, and will be denoted by $d(f,g)$. We say that an $n$-variable Boolean function $f$ is balanced if $\text{wt}(f) = 2^{n-1}$.

Let $f \in B_n$. The nonlinearity of $f$ is its distance from the set of all $n$-variable affine functions, i.e.,

$$\text{nl}(f) = \min_{g \in A_n} d(f,g).$$
The nonlinearity of an \( n \)-variable Boolean function is bounded above by \( 2^{n-1} - 2^{n/2 - 1} \), and a function is said to be bent if it achieves this bound. Clearly, bent functions exist only for even \( n \) and it is known that the algebraic degree of a bent function is bounded above by \( \frac{n}{2} \) \([13]\).

A Boolean function \( f \in B_n \) is called \( k \)-normal (respectively \( k \)-weakly-normal) if there exist a \( k \)-dimensional constant (respectively affine) flat for \( f \). If \( k = \lfloor \frac{n}{2} \rfloor \), \( f \) is simply called a normal (respectively weakly-normal) function.

For any \( f \in B_n \), a nonzero function \( g \in B_n \) is called an annihilator of \( f \) if \( fg \) is null, and the algebraic immunity of \( f \), denoted by \( AI(f) \), is the minimum value of \( d \) such that \( f \) or \( f + 1 \) admits an annihilator of degree \( d \) \([25]\). It is known that the algebraic immunity of an \( n \)-variable Boolean function is bounded above by \( \lceil \frac{n}{2} \rceil \) \([12]\).

To resist algebraic attacks, a Boolean function \( f \) should have a high algebraic immunity, which implies that the nonlinearity of \( f \) is also not very low, since, according to Lobanov’s bound \([23]\):

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

To resist fast algebraic attacks, a high algebraic immunity is not sufficient. If we can find \( g \) of low degree and \( h \) of algebraic degree not much larger than \( n/2 \) such that \( fg = h \), then \( f \) is considered to be weak against fast algebraic attacks \([11, 17]\). The higher order nonlinearities of functions with high (fast) algebraic immunity is also not very low \([6, 26, 37]\).

The Walsh transform of a given function \( f \in B_n \) is the integer-valued function over \( \mathbb{F}_2^n \) defined by

\[
W_f(u) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) + u \cdot x},
\]

where \( u \in \mathbb{F}_2^n \) and \( u \cdot x \) is an inner product, for instance, \( u \cdot x = u_1 x_1 + u_2 x_2 + \cdots + u_n x_n \), where \( u = (u_1, \ldots, u_n), x = (x_1, \ldots, x_n) \). A function \( f \) is said to be resilient of order \( r \) if \( W_f(u) = 0 \), for \( 0 \leq wt(u) \leq r \). It is easy to see that a Boolean function \( f \) is balanced if and only if \( W_f(0) = 0 \). Moreover, the nonlinearity of \( f \) can be determined by

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

The autocorrelation function of \( f \in B_n \) is defined by

\[
C_f(u) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) + f(x+u)}.
\]

Also, \( f \) satisfies the strict avalanche criterion (SAC) if \( C_f(a) = 0 \), for \( wt(a) = 1 \).

We say that \( f, g \in B_n \) are affine equivalent if there exists an \( n \times n \) invertible matrix \( A \) over the finite field \( \mathbb{F}_2 \) and vectors \( b, c \in \mathbb{F}_2^n \) such that \( g(x) = f(Ax + b) + c \cdot x \). If \( f \) is a Boolean function in \( n \) variables and \( H \) is a flat (a coset of some vector subspace) in \( \mathbb{F}_2^n \), we let \( f|_H \) to be the (restriction) function \( f|_H : H \to \mathbb{F}_2 \) by \( f|_H(x) = f(x) \), for \( x \in H \).
3 The construction and its cryptographic properties

We next introduce a construction based on balanced functions \( f_i \in \mathcal{B}_{n-2}, i = 1, 2 \) (|| denotes concatenation).

**Construction 1.** For \( \{i, j\} = \{1, 2\} \), we define the functions \( f \) on \( \mathbb{F}_2^n \):

\[
\begin{align*}
f &= f_i || f_j || f_j = x_{n-1} x_n + x_{n-1} (f_i + f_j) + f_i, \\
&\text{which is affine equivalent to } f_i || f_j || f_j: f_i || f_j || f_j: f_i || f_j || f_j; f_i || f_j || f_j; f_i \parallel f_i \parallel f_j.
\end{align*}
\]

\[
\begin{align*}
f &= f_i || f_j || f_j \bar{f}_i = x_{n-1} x_n + x_{n-1} (f_i + f_j) + x_n (f_i + f_j) + f_i, \\
&\text{which is affine equivalent to } f_i || f_j || f_j: f_i || f_j || f_j: f_i || f_j || f_j; f_i || f_j || f_j.
\end{align*}
\]

We could have defined \( f = f_i || f_j || \bar{f}_j = x_{n-1} (f_i + f_j) + x_n + f_i \) to also fulfil resiliency, but, unfortunately, a few of the other properties get slightly weaker.

Some of the cryptographic properties, like algebraic immunity, or strict avalanche criterion are not (fully) affine invariants, so we prefer to list all concatenations for completeness, but we shall prove our results motivating whenever necessary, the relevant differences between the various classes of functions.

Variations of this construction have appeared in literature and some of their properties have been investigated. Most notably, bent, resiliency and the normality properties of a concatenation (based upon bent functions) were looked at in \([5, 8, 9]\), and the normality of \( f_i || f_j || f_j || f_j \) for arbitrary \( f_i \) with \( i = 1, 2 \) is addressed in \([16]\) (see also \([29]\) for other properties). We will also prove the normality result later on for all these classes. We mention also the paper of Pasalic \([27]\), which introduces the notion of high degree product (HDP) to measure the ability of Boolean functions to be resistant to fast algebraic attacks: \( f \in \mathcal{B}_n \) satisfies the HDP of order \( n \) if any non-annihilating function \( g \) of degree \( e \) with \( 1 \leq e \leq \lceil n/2 \rceil - 1 \), the degree \( d = \deg(gf) \) satisfies \( e + d \geq n \).

Unfortunately, the construction of \([27]\) based upon a four function concatenation does not always produce almost optimal HDP functions, and this was observed in \([39]\).

We next generalize a known lemma that relates the Walsh–Hadamard coefficients of \( g \) (in some dimension) to the Walsh–Hadamard coefficients of its \( 2^r \) \((r \geq 1)\) concatenation parts.

**Lemma 1** If \( g(x, x_{n+1}, \ldots, x_{n+r}) = f_1(x) || f_2(x) || \cdots || f_{2^r}(x) = 2^r \bigwedge_{i=1}^{2^r} f_i(x) \), then

\[
\begin{align*}
W_g(u, u_{n+1}, \ldots, u_{n+r}) &= W_{f_1}(u) + (-1)^{u_{n+1}} W_{f_2}(u) + \cdots + (-1)^{u_{n+r}} W_{f_{2^r}}(u) \\
&= \sum_{k=1}^{2^r} (-1)^{a(k) \cdot \mathbf{u}} W_{f_k}(u),
\end{align*}
\]

where \( r \in \mathbb{N} \), \( a(k) \) is the \( k \)th lexicographically ordered vector in \( \mathbb{F}_2^r \), and \( \mathbf{u}' = (u_{n+1}, \ldots, u_{n+r}) \).
Proof. We show our result by induction on $r$. If $r = 1$, we compute

$$\begin{align*}
W_g(u, u_{n+1}) &= \sum_{(x, x_{n+1}) \in F_2^n} (-1)^g(x, x_{n+1}) + u \cdot x + u_{n+1} x_{n+1} \\
&= \sum_{x \in F_2^n} (-1)^{g_1(x) + u \cdot x} + (-1)^{u_{n+1}} \sum_{x \in F_2^n} (-1)^{g_2(x) + u \cdot x} \\
&= W_{g_1}(u) + (-1)^{u_{n+1}} W_{g_2}(u).
\end{align*}$$

In general, assuming that $W_g(u, u_{n+1}, \ldots, u_{n+r}) = \sum_{k=1}^{2^n} (1)^{a(k) \cdot u'} W_{f_k}(u)$, then for $g''(x, x_{n+1}, \ldots, x_{n+r+1}) = f_1(x) \| f_2(x) \| \cdots \| f_{2^r+1}(x) = g' = \frac{1}{i=1} f_i(x)$, where $g' = f_{2^r+1}(x) \| f_{2^r+2}(x) \| \cdots \| f_{2^{r+1}}(x)$, we have

$$\begin{align*}
W_{g''}(u, u_{n+1}, \ldots, u_{n+r+1}) &= W_g(u, u_{n+1}, \ldots, u_{n+r}) + (-1)^{u_{n+r+1}} W_{g'}(u, u_{n+1}, \ldots, u_{n+r}) \\
&= W_{f_1}(u) + (-1)^{u_{n+1}} W_{f_2}(u) + \cdots + (-1)^{u_{n+r+1}} W_{f_{2^r+1}}(u) \\
&= \sum_{k=1}^{2^{r+1}} (1)^{a(k) \cdot u'} W_{f_k}(u),
\end{align*}$$

which shows our claim. 

Lemma 2 (Proposition 1 of [7]) Let $g_1$, $g_2$ be two Boolean functions in the variables $x_1, \ldots, x_n$ with $AI(g_1) = d_1, AI(g_2) = d_2$, and let $g = (1 + x_{n+1}) g_1 + x_{n+2} g_2 \in B_{n+2}$. Then, the following hold:

1. If $d_1 \neq d_2$, then $AI(g) = \min\{d_1, d_2\} + 1$.
2. If $d_1 = d_2 (=: d)$, then $d \leq AI(g) \leq d + 1$. Further, $AI(g) = d$ if and only if there exists $f_1, f_2 \in B_n$ of algebraic degrees $d$ that either both annihilate $g_1, g_2$, or both annihilate $g_1, g_2$, and $\deg(f_1 + f_2) \leq d - 1$.

For our next result, we let $f_1 \in B_{n-2}$ in Construction 1 be any balanced function and $f_2(x) = f_1(Ax + b)$, where $A$ is an $(n-2)$ by $(n-2)$ invertible matrix over the finite field $F_2$ and $b$ is an $(n-2)$ dimensional vector over $F_2$. Clearly, $f_1$ and $f_2$ are affine equivalent, $\deg(f_1) = \deg(f_2)$, $AI(f_1) = AI(f_2)$ and $nl(f_1) = nl(f_2)$.

Theorem 3 Let $f$ be given by Construction 1, with $f_1, f_2$ nonconstant and affine equivalent. Then $f$ is balanced, $\deg(f) = \max\{\deg(f_1), \deg(f_1 + f_2) + 1\}$, $AI(f_1) + 2 \geq AI(f) \geq \min\{AI(f_1||f_2), AI(f_1||f_2)\} \geq AI(f_1)$. Moreover, $nl(f) = 2^{n-2} + 2nl(f_1)$.

Proof. We take $f = f_1||f_2||f_1||f_2$ as an example (the other cases equivalent to this one in Construction 1 are similar). Clearly, $\deg(f) = \deg(f_1||f_2) = \max\{\deg(f_1), \deg(f_1 + f_2) + 1\}$. Since $f_1||f_2$ is affine equivalent to $f_1||f_2$ (precisely, $(f_1||f_2)(x) = (f_1||f_2)(x) + x_n$ we have $|AI(f_1||f_2) - AI(f_1||f_2)| \leq 1$ (by
If \( \mathcal{AI}(f_1 \| f_2) = \mathcal{AI}(f_1 \| \bar{f}_2) \), by Lemma 2, \( \mathcal{AI}(f) \geq \mathcal{AI}(f_1 \| f_2) \geq \mathcal{AI}(f_1) \). If \( |\mathcal{AI}(f_1 \| f_2) - \mathcal{AI}(f_1 \| \bar{f}_2)| = 1 \), then Lemma 2 shows that \( \mathcal{AI}(f) = \min\{d, d + 1\} + 1 = d + 1 \), where \( \min\{\mathcal{AI}(f_1 \| f_2), \mathcal{AI}(f_1 \| \bar{f}_2)\} = d \).

If \( f = f_1 || f_2 || \bar{f}_1 \) (the other cases equivalent to this one in Construction 1 are similar), then \( \deg(f) = \max\{\deg(f_1 || f_2), \deg((f_1 + \bar{f}_2) || (f_2 + f_1)) + 1\} \), and since \( \deg((f_1 + f_2) || (f_2 + f_1)) = \max\{\deg(f_1 + f_2), \deg(f_1 + f_2 + f_2 + f_1) + 1\} = \max\{\deg(f_1 + f_2), 1\} \leq \deg(f_1) \), we get that \( \deg(f) = \deg(f_1 || f_2) = \max\{\deg(f_1), \deg(f_1 + f_2) + 1\} \), in this case, as well. The algebraic immunity computation does not change in this case.

To find the nonlinearity, we first consider \( f = f_1 || f_2 || f_1 || f_2 || \bar{f}_1 \) of Construction 1 (the others are similar). Using Lemma 1, we obtain

\[
W_f(u, u_{n-1}, u_n) = W_{f_1}(u) + (-1)^{u_n-1} W_{f_2}(u) + (-1)^{u_n} W_{f_1}(u) + (-1)^{u_n-1+u_n} W_{f_2}(u) = (1 + (-1)^{u_n}) W_{f_1}(u) + (-1)^{u_n-1} (1 - (-1)^{u_n}) W_{f_2}(u).
\]

Thus, \( W_f(u, u_{n-1}, 0) = 2W_{f_1}(u) \) and \( W_f(u, u_{n-1}, 1) = 2(-1)^{u_n-1} W_{f_2}(u) \). It follows that

\[
\max_{(u, u_{n-1}, u_n) \in \mathbb{F}_2^n} |W_f(u, u_{n-1}, u_n)| = 2 \max_{u \in \mathbb{F}_2^{n-1}} |W_{f_1}(u)| = 2^{n-1} - 4nl(f_1).
\]

Therefore, \( nl(f) = 2^{n-2} + 2nl(f_1) \). \( \square \)

**Remark 4** In most cases, \( \deg(f_1(Ax + b) + f_1) = \deg(f_1) \). That is, \( \deg(f) = \deg(f_1) + 1 \). As can be seen from Proposition 1 of [7], in many cases, \( \mathcal{AI}(f_1 || f_2) = \mathcal{AI}(f_1) + 1 \), and so, \( \mathcal{AI}(f) \geq \mathcal{AI}(f_1) + 1 \). Further, \( nl(f) \) is obviously higher than \( nl(f_1) \). In fact, the fast correlation attack on \( f \) has an on-line complexity proportional to \( \left( \frac{1}{2} \right)^2 \), where \( \epsilon = \frac{1}{2} - \frac{nl(f_1)}{2^{n-1}} \) is the so-called bias [24]. We have

\[
\frac{1}{2} - \frac{nl(f)}{2^n} = \frac{1}{2} - \frac{1}{4} - \frac{nl(f_1)}{2^{n-1}} = \frac{1}{2} \left( \frac{1}{2} - \frac{nl(f_1)}{2^{n-2}} \right),
\]

that is, the bias of \( f \) is half of the bias of \( f_1 \).

**Remark 5** Using Proposition 1 of [27], we infer that if one takes \( f_1, f_2 \in \mathcal{B}_n \) \( (n \text{ even}) \) of maximum \( \mathcal{AI} \), with the property that for any function \( g \) of algebraic degree \( 1 \leq e \leq \lceil n/2 \rceil - 1 \), we have \( \deg(f_1g) = d \geq \mathcal{AI}(f_1) \) and \( e + d \geq n \), then \( f = f_1 || f_2 || f_1 || f_2 \) has maximum \( \mathcal{AI} \). Such functions \( f_1 \) are called perfect algebraic immune (\( \mathcal{PAI} \)). The authors of [22] showed (in their Theorem 7) that if \( f_1 \) is a balanced \( \mathcal{PAI} \) then \( n = 2^k + 1 \), for some \( k \); if \( f_1 \) is unbalanced, then \( n = 2^k \), for some \( k \).

Next, we are concerned with normality issues of our construction. References [8, 16] contain secondary constructions of normal, or non-normal functions based upon some of the functions of Construction 1, namely \( f_1 || f_2 || f_2 || f_1 \), where \( f_i \) are bent or have some normality properties.
Theorem 6 Let \( f_1, f_j \in B_{n-2} \). If \( f_i \) or \( f_j \), whichever is not complemented in Construction 1, is \( k \)-normal, then the functions \( f \) of Construction 1 are at least \((k + 1)\)-normal.

Proof. Due to the affine equivalence to \( f_i \), \( f_j \) is \( k \)-normal. If \( f_i \) is invariant, say 0 on a \( k \)-dimensional flat, then \( f_i \) is constant and equal to 1 on the same flat, which shows that \( f_i \) is \( k \)-normal. We prove the case \( f = f_i \| f_j \| f_1 \| f_j \) only, since other cases can be shown by similar arguments. We show the existence of a \((k + 1)\)-dimensional affine subspace where \( f(x) \) is a constant. Let \( z_1, \ldots, z_k \in \mathbb{F}_2 \) be \( k \) distinct, linearly independent vectors in \( \mathbb{F}_2^{n-2} \). Let \( d = (d_1, d_2, \ldots, d_{n-2}) \) be a vector in \( \mathbb{F}_2^{n-2} \), and \( a_i \in \mathbb{F}_2 \) be for \( 1 \leq i \leq k \). We define a \( k \)-dimensional flat \( G = \{ x \in \mathbb{F}_2^{n-2} \mid x = a_1z_1 + a_2z_2 + \cdots + a_kz_k + d, a_i = \mathbb{F}_2, 1 \leq i \leq k \} \) such that \( f_i \mid G = 0 \). We now construct a \((k + 1)\)-dimensional flat in the following way. Let \( z_l = (z_{l1}, z_{l2}, \ldots, z_{(n-2)}) \) where \( 1 \leq l \leq k \). We set \( z'_l = (z_{l1}, z_{l2}, \ldots, z_{(n-2)}, 0, 0) \), \( z'_{k+1} = (0, \ldots, 0, 1) \), and \( d' = (d_1, d_2, \ldots, d_{n-2}, 0, 0) \) where \( z'_{k+1} \in \mathbb{F}_2^2 \). Then \( G' = \{ x' \in \mathbb{F}_2^2 \mid x' = a_1z'_1 + a_2z'_2 + \cdots + a_{k+1}z'_{k+1} + d' \} \), \( a_l = \mathbb{F}_2, 1 \leq l \leq k+1 \). If a vector \( x' \in G' \) with \( a_{k+1} = 0 \), then \( f \) equals the first \( f_i \) in the construction. If a vector \( x' \in G' \) with \( a_{k+1} = 1 \), then \( f \) has the same value as the third \( f_i \) in the construction. Therefore, \( G' \) is a \((k + 1)\)-dimensional flat such that \( f \mid G' = 0 \). □

Generally, it is difficult to establish a proper limit to the normality of a function. Let \( f_i \) or \( f_j \), whichever is not complemented in Construction 1, be \( k \)-normal but not \((k + 1)\)-normal, and we show that the function \( f \) of Construction 1 cannot have a constant function value on the \( k+2 \)-dimensional flat \( H = \{ e_1z_1 + e_2z_2 + \cdots + e_{k+2}z_{k+2} \} \), where \( d = (y_1, \ldots, y_n) \) is a fixed vector in \( \mathbb{F}_2^2 \) and \( e_m = (x_1, \ldots, x_n) \) is an elementary vector such that \( x_j = 1 \) if and only if \( j = i_m \) with \( 1 \leq i_m \leq n \). We assume \( f = f_i \| f_j \| f_1 \| f_j \) since the other cases can be shown by similar arguments. Let us also assume that \( H \) exists such that \( f \mid H \) is constant. We observe that \( y_{i_m} \) is irrelevant (whether it is 0 or 1) due to \( e_{i_m} \), so we let \( d \) with \( y_{i_1} = \ldots = y_{i_{k+2}} = 0 \). To illustrate better, we rewrite the restriction of our function to \( H \) as follows:

\[
\begin{align*}
\| f(x) \|_H &= (\bar{x}_{n-1}f_i \oplus x_{n-1}f_j) \| (\bar{x}_{n-1}f_i \oplus x_{n-1}f_j) \|_H \\
&= \bar{x}_n(\bar{x}_{n-1}f_i \oplus x_{n-1}f_j) \oplus x_n(\bar{x}_{n-1}f_i \oplus x_{n-1}f_j) \|_H \\
&= \bar{x}_{n-1}(\bar{x}_n f_i \oplus x_n f_j) \oplus x_{n-1}(\bar{x}_n f_i \oplus x_n f_j) \|_H \\
&= f_i \oplus x_{n-1}f_j \oplus x_n f_j \oplus x_n f_j \|_H.
\end{align*}
\]

Without loss of generality, we assume \( f(x) = 0 \) for all \( x = (x_1, \ldots, x_n) \in H \), and we examine the following cases, depending upon the values of \( x_{n-1} \) and \( x_n \).

Case 1: \( n = 1 \), \( n \notin \{i_1, i_2, \ldots, i_{k+2}\} \). Then \( x_{n-1} = d_{n-1} \), and \( d_n = x_n \). We observe that for all possible values for \( x_{n-1} \) and \( x_n \), \( f \mid H \) is one of the functions, \( f_i \), \( f_j \), or \( f \). Since each function is only \( k \)-normal, there exists at least one \( x \in H \) such that \( f(x) \mid H = 1 \), which is a contradiction.

Case 2: \( n-1 \notin \{i_1, i_2, \ldots, i_{k+2}\} \) and \( x_n \in \{i_1, i_2, \ldots, i_{k+2}\} \). Then \( x_{n-1} = d_{n-1} \). If \( x_{n-1} = 0 \), then regardless of the value of \( x_n \), \( f \mid H \) equals the function \( f_i \). We note that we can only increase the normality to \( k + 1 \) using \( x_n \), since \( f_i \) is \( k \)-normal. Therefore, there exists at least one \( x \in H \) such that \( f(x) \mid H = 1 \), which
which is balanced (regardless of whether \( f \) with \( x_n = 0 \) or \( \bar{f}_j \) with \( x_n = 1 \). Clearly, \( f \) is at most \( k \)-normal, since \( f_j = f_j \oplus 1 \). So, there exists at least one \( x \in H \) such that \( f(x)\|_H = 1 \), which is a contradiction.

**Case 3:** \( n \notin \{i_1, i_2, \ldots, i_{k+2}\} \) and \( x_{n-1} \in \{i_1, i_2, \ldots, i_{k+2}\} \). Then \( d_n = x_n \). If \( x_n = 0 \), then \( f \|_H \) equals the function, \( f_i \| f_j \). Also, if \( x_n = 1 \), then \( f \|_H \) equals the function \( f_i \| f_j \). In both instances, we can only increase the normality to \( k + 1 \), since \( f_i \), \( f_j \) and \( f_k \) are \( k \)-normal.

**Case 4:** \( x_{n-1}, x_n \in \{i_1, i_2, \ldots, i_{k+2}\} \). In this case \( f \|_H \) equals \( f_i \| f_j \| f_k \|_H \), and any two vectors \( x', x'' \in H \) in the forms of \( x' = (a_1, \ldots, a_{n-2}, 1, 0) \) and \( x'' = (b_1, \ldots, b_{n-2}, 1, 1) \) with \( a_i, b_i \in \mathbb{F}_2, 1 \leq i \leq n - 2 \) have opposite function values. Therefore, we have a contradiction.

Under what conditions the functions of Construction 1 is exactly \((k + 2)\)-normal remains an open problem.

**Theorem 7** If the base functions \( f_1 \) and \( f_2 \) in Construction 1 satisfy the strict avalanche criterion, then \( f \) satisfies the strict avalanche criterion.

**Proof.** For every vector \( y \in \mathbb{F}_2^n \), write \( y = (y', y_{n-1}, y_n) \) with \( y' \in \mathbb{F}_2^{n-2} \). We shall show the claim in the case \( f = f_1 \| f_2 \| f_1 \| f_2 \), as all the other possibilities are similar. Let \( a \in \mathbb{F}_2^n \) of weight \( wt(a) = 1 \). We consider three cases.

**Case 1.** Take \( a = (0, \ldots, 0, 1) \). Then

\[
\begin{align*}
f(x) + f(x + a) &= (f_1 \| f_2)(x', x_{n-1}) \bar{x}_n + (f_1 \| \bar{f}_2)(x', x_{n-1}) x_n \\
&\quad + (f_1 \| f_2)(x', x_{n-1}) x_n + (f_1 \| \bar{f}_2)(x', x_{n-1}) \bar{x}_n \\
&= (f_1 \| f_2)(x', x_{n-1}) + (f_1 \| \bar{f}_2)(x', x_{n-1}) = 0_{2^{n-1}} \| 1_{2^{n-1}},
\end{align*}
\]

hence balanced.

**Case 2.** Take \( a = (0, \ldots, 1, 0) \). Then

\[
\begin{align*}
f(x) + f(x + a) &= (f_1 \| f_2)(x', x_{n-1}) \bar{x}_n + (f_1 \| \bar{f}_2)(x', x_{n-1}) x_n \\
&\quad + (f_1 \| f_2)(x', x_{n-1}) x_n + (f_1 \| \bar{f}_2)(x', x_{n-1}) \bar{x}_n \\
&= f_1(x') x_{n-1} \bar{x}_n + f_2(x') \bar{x}_{n-1} x_n + f_1(x') x_{n-1} x_n + f_2(x') \bar{x}_{n-1} x_n \\
&\quad + f_1(x') \bar{x}_{n-1} \bar{x}_n + f_2(x') x_{n-1} x_n + f_1(x') \bar{x}_{n-1} x_n + f_2(x') x_{n-1} x_n \\
&= f_1(x') \bar{x}_n + f_2(x') \bar{x}_n + f_1(x') x_n + f_2(x') x_n = f_1(x') + f_2(x') + x_n,
\end{align*}
\]

which is balanced (regardless of whether \( f_1 + f_2 \) is balanced).
Case 3. Take \(a = (a', 0, 0)\), with \(wt(a') = 1\). Write \(x' + a' = x''\). Then,

\[
f(x) + f(x + a) = (f_1 || f_2)(x', x_{n-1}) \bar{x}_n + (f_1 || \bar{f}_2)(x', x_{n-1}) x_n
\]

\[
+ (f_1 || f_2)(x'', x_{n-1}) \bar{x}_n + (f_1 || \bar{f}_2)(x'', x_{n-1}) x_n
\]

\[
f_1(x') \bar{x}_{n-1} \bar{x}_n + f_2(x') x_{n-1} \bar{x}_n + f_1(x') \bar{x}_{n-1} x_n + \bar{f}_2(x') x_{n-1} x_n
\]

\[
+ f_1(x'') \bar{x}_{n-1} \bar{x}_n + f_2(x'') x_{n-1} \bar{x}_n + f_1(x'') \bar{x}_{n-1} x_n + \bar{f}_2(x'') x_{n-1} x_n
\]

\[
= (f_1(x') + f_1(x'')) \bar{x}_{n-1} \bar{x}_n + (f_2(x') + f_2(x'')) x_{n-1} \bar{x}_n
\]

\[
+ (f_1(x') + f_1(x'')) \bar{x}_{n-1} x_n + (f_2(x') + f_2(x'')) x_{n-1} x_n
\]

\[
= (f_1(x') + f_1(x'')) \bar{x}_{n-1} + (f_2(x') + f_2(x'')) x_{n-1},
\]

which is balanced, since \(f_1, f_2\) satisfy the strict avalanche criterion, and so, both \(f_1(x') + f_1(x' + a')\) and \(f_2(x') + f_2(x' + a')\) are balanced.

4 Two particular cases, based on the HWBF and Carlet-Feng function

In 2002, Krause [18] introduced another attack using Binary Decision Diagram (BDD). Later research [19, 33] showed the effectiveness of BDD-based attacks on stream ciphers (albeit, they generally require a large amount of memory). Krause notes that one of the effective ways to disrupt BDD-based attacks is for the Boolean function combiner of the stream cipher to have a robust BDD.

There are many constructions of Boolean functions with high algebraic immunity [1, 7, 9, 10, 14, 15, 20, 21, 28, 30–32, 34–36, 38, 40]. However, none of these papers (except for [38]) took BDD-based attacks into consideration. Interestingly, Bryant (see [2, 3]) showed that the Hidden Weighted Bit Function (HWBF) has exponential size BDD.

Below, we give more exact results on the cryptographic properties of functions in our construction if the base functions are variations of the hidden weighted bit function (HWBF), which is defined by

\[
h_n(x) = \begin{cases} 
0 & \text{if } x = 0 \\
\text{wt}(x) & \text{otherwise}.
\end{cases}
\]

It is known [38] that \(h_n \in B_n\) is balanced, with the optimum algebraic degree, satisfying the strict avalanche criterion, \(\max_{u \in F_2^n} |W_{h_n}(u)| = 4\left(\frac{n-2}{2}\right)\) and algebraic immunity \(AI(h_n) \geq \left\lfloor \frac{n}{2} \right\rfloor + 1\).

Let \(\phi\) be the left-rotation symmetric operation on vectors of arbitrary dimension, say \(\phi(x_1, x_2, \ldots, x_n) = (x_2, \ldots, x_n, x_1)\). It was shown previously in [38] that the hidden weighted bit (HWBF) function is a concatenation which can be iterated (write \(x = (x_2, \ldots, x_{n-2}) \in F_2^{n-3}\), as shown in the next formula

\[
h_n(x_1, x, x_{n-1}, x_n) = h_{n-1}(x_1, x, x_{n-1}) || (h_{n-1} \circ \phi)(x_1, x, x_{n-1})
\]

\[
= h_{n-2}(x_1, x) || (h_{n-2} \circ \phi)(x_1, x) || h_{n-2}(x, x_{n-1}) || (h_{n-2} \circ \phi)(x, x_{n-1}) \ldots
\]

(1)
Theorem 8  Let $n \geq 3$ and $f_1||f_2 = h_{n-1}$ be an $(n-1)$-variables HWBF. Then, all of the functions $f$ from Construction 1 are balanced of degree $\max\{n-2,2\}$, have nonlinearity $nl(f) = 2^n - 4(\frac{n-4}{(n-4)/2})$, and have algebraic immunity $\geq \frac{n+2}{3}$.

Proof. Certainly, all functions in Construction 1 are balanced. Furthermore, for any concatenation $g_1||g_2$, the degree $\deg(g_1||g_2) = \max\{\deg(g_1), \deg(g_1+g_2)+1\}$.
Thus, since $\deg(f_1||f_2) = \deg(h_{n-1}) = n - 2$, we obtain
\[
\deg(f_1||f_2||f_1||\tilde{f}_2) = \max\{\deg(f_1||f_2), \deg((f_1||f_2) + (f_1||\tilde{f}_2)) + 1\} = \max\{n - 2, \deg(0_{2^{n-2}12^{n-2}}) + 1\} = \max\{n - 2, 2\}
\]
(we write $0_s$, or $1_s$, for the corresponding bit repeated $s$ times).

Next, we will do the computation for only one case, say $f = f_1||f_2||f_1||\tilde{f}_2$.

We will show that $\max_{w \in F_2^{n-2}} |W_f(w)| = 8(\frac{n-4}{n-2})$.

Using Lemma 1, with $g_1 = h_{n-1} = f_1||f_2$, $g_2 = f_1||\tilde{f}_2$, $f_1 = h_{n-2}$, and $f_2 = h_{n-2} \circ \phi$, as in the proof of Theorem 3 we obtain
\[
W_f(u, u_{n-1}, u_{n}) = (1 + (-1)^u)v W_f_1(u) + (-1)^v(1 - (-1)^u)v W_f_2(u).
\]
Thus, $W_f(u, u_{n-1}, 0) = 2W_f_1(u)$ and $W_f(u, u_{n-1}, 1) = 2(-1)^u W_f_2(u)$. Since $f_1(u) = h_{n-2}(u)$ and $f_2(u) = h_{n-2}(\phi(u))$ and $\max_{u \in F_2^2} |W_{h_{n-2}}(u)| = 4(\frac{n-4}{n-2})$, it follows that
\[
\max_{(u, u_{n-1}, u_{n}) \in F_2^n} |W_f(u, u_{n-1}, u_{n})| = 2 \max\{\max_{u \in F_2^{n-2}} |W_{h_{n-2}}(u)|, \max_{u \in F_2^{n-2}} |W_{h_{n-2}}(\phi(u))|\} = 8(\frac{n-4}{n-2}).
\]

We obtain that the nonlinearity of the functions in Construction 1 is $nl(f) = 2^n - 4(\frac{n-4}{n-2})$.

We now deal with the computation of the algebraic immunity for the considered functions. By Theorem 4 of [38], $AI(h_n) =: d_n \geq \frac{n}{2} + 1$. We observe that $h'_n(x_1, x_2, \ldots, x_n) = h_n(x_2, x_3, \ldots, x_n, x_1)$, and certainly $AI(h_n) = AI(h'_n)$. By the definition of algebraic immunity, we observe that $AI(g) = AI(\overline{g})$ for any Boolean function $g$, and also that $AI(f_i||f_j) = AI(f_i||f_j)$, and $AI(f_i||f_j) = AI(f_i||f_j)$, $i = 1, 2$. So without loss of generality, we will only consider the case of $f = f_1||f_2||f_1||\tilde{f}_2$.

Let $g = g_1||g_2||k_1||k_2 \neq 0$ be an annihilator of $f$. Thus, $g_1, k_1$ are both annihilators of $f_1$; and, $g_2, k_2$ are annihilators of $f_2$, respectively, $\tilde{f}_2$, not all zero.

First, since $g_1||g_2$ is an annihilator of $f_1||f_2 = h_{n-1}$, it follows that $\deg(g_1||g_2) = 0$, if both $g_1 = g_2 = 0$, or $\deg(g_1||g_2) \geq d_{n-1}$. Also, observe that $\deg(g_1 + k_1)$ is
Table 1. Algebraic immunity and nonlinearity of the HWBF-based $f$ and the HWBF $h$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$AI(f)$</th>
<th>$AI(h)$</th>
<th>$nl(f)$</th>
<th>$nl(h)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>3</td>
<td>3</td>
<td>52</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
<td>104</td>
<td>88</td>
</tr>
<tr>
<td>9</td>
<td>4</td>
<td>4</td>
<td>216</td>
<td>186</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>5</td>
<td>432</td>
<td>372</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>5</td>
<td>884</td>
<td>772</td>
</tr>
<tr>
<td>12</td>
<td>5</td>
<td>5</td>
<td>1768</td>
<td>1544</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>5</td>
<td>3592</td>
<td>3172</td>
</tr>
<tr>
<td>14</td>
<td>6</td>
<td>5</td>
<td>7184</td>
<td>6344</td>
</tr>
<tr>
<td>15</td>
<td>6</td>
<td>6</td>
<td>14536</td>
<td>12952</td>
</tr>
</tbody>
</table>

Table 2. Behavior of the HWBF-based function $f$ against Fast Algebraic Attacks

<table>
<thead>
<tr>
<th>$n$</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(d,e)$</td>
<td>(1,3)</td>
<td>(1,5)</td>
<td>(1,5)</td>
<td>(1,7)</td>
<td>(1,9)</td>
<td>(1,9)</td>
<td>(2,4)</td>
</tr>
<tr>
<td></td>
<td>(2,4)</td>
<td>(2,4)</td>
<td>(2,5)</td>
<td>(2,6)</td>
<td>(2,8)</td>
<td>(2,8)</td>
<td>(3,3)</td>
</tr>
<tr>
<td></td>
<td>(3,4)</td>
<td>(3,4)</td>
<td>(3,5)</td>
<td>(3,5)</td>
<td>(3,6)</td>
<td>(3,6)</td>
<td>(4,5)</td>
</tr>
<tr>
<td></td>
<td>(4,5)</td>
<td>(4,6)</td>
<td>(4,6)</td>
<td>(4,6)</td>
<td>(4,6)</td>
<td>(4,6)</td>
<td>(5,6)</td>
</tr>
</tbody>
</table>

either 0, if $g_1 = k_1$, or $\geq d_{n-1}$ (since $g_1 + k_1$ is an annihilator of $f_1$). Now, the degree of the concatenation $g = g_1 || g_2 || k_1 || k_2$ is

$$\deg(g) = \max\{\deg(g_1 || g_2), \deg((g_1 + k_1) || (g_2 + k_2)) + 1\}.$$ 

Next, $\deg(g_1 || g_2) = \max\{\deg(g_1), \deg(g_1 + g_2) + 1\}$, and $\deg((g_1 + k_1) || (g_2 + k_2)) = \max\{\deg(g_1 + k_1), \deg(g_1 + g_2 + k_1 + k_2) + 1\}$.

It is rather obvious that the worst case is when $g_1 = k_1, g_2 = k_2$. Then, $\deg(g) = \max\{\deg(g_1 || g_2), 1\} \geq \left\lceil \frac{n-1}{3} \right\rceil + 1 = \left\lceil \frac{n+2}{3} \right\rceil$, which proves the first claims.

Let $f = f_1 || f_2 || f_1 || f_2$ with $f_1 f_2 = h_{n-1}$, the HWBF. In Table 1, one can find the algebraic immunity and nonlinearity of $f$, compared to the same parameters for the HWBF $h$. Let $fg = h$, $\deg(g) = d$ and $\deg(h) = e$. In Table 2, we give the lowest possible values of $(d,e)$, as needed for the fast algebraic attack.

There are other cases where our construction is quite strong, improving upon the parameters of other existing constructions. Let $CF \in B_n$ be the Carlet–Feng function whose support is $\{0, 1, \alpha, \ldots, \alpha^{n-1-2}\}$, where $\alpha$ is a primitive element of the field $\mathbb{F}_{2^n}$. We now consider our construction with the base functions that are variations of $CF$, and give some experimental results.

Let $f_1 \in B_{10}$ be the Carlet–Feng function with the primitive polynomial

$$2^{10} + 2^{6} + 2^{5} + 2^{3} + 2 + 1$$
Table 3. Algebraic immunity and nonlinearity of the CF-based function $f$ and the function $TCT$ introduced by [35]

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\text{AI}(f)$</th>
<th>$\text{AI}(TCT)$</th>
<th>$\text{nl}(f)$</th>
<th>$\text{nl}(TCT)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>4</td>
<td>4</td>
<td>112</td>
<td>108</td>
</tr>
<tr>
<td>10</td>
<td>5</td>
<td>5</td>
<td>480</td>
<td>476</td>
</tr>
<tr>
<td>12</td>
<td>6</td>
<td>6</td>
<td>1992</td>
<td>1982</td>
</tr>
<tr>
<td>14</td>
<td>7</td>
<td>7</td>
<td>8076</td>
<td>8028</td>
</tr>
<tr>
<td>16</td>
<td>8</td>
<td>8</td>
<td>32532</td>
<td>32508</td>
</tr>
</tbody>
</table>

and $f_2(x) = f_1(Ax)$, where

$$A = (e_1, e_2, e_3, e_4, e_5, e_{10}, e_6, e_7, e_8, e_9)$$

and $e_i \in \mathbb{F}_2^{10}$ is the unit column vector with 1 on the $i$-th position. Let $f = f_1 || f_2 || f_2 \in B_{12}$. Then $\text{AI}(f) = 6$ and $\text{nl}(f) = 1992$. The nonlinearity of the 12-variable Carlet-Feng function discussed by [35, 36] is only 1970. Also, the balanced function $TCT : \mathbb{F}_{2k} \times \mathbb{F}_{2k} \to \mathbb{F}_2$ constructed by [35] which is based on the Carlet–Feng function has the nonlinearity 1982, when $2k = 12$.

Clearly, our function $f$ has optimum algebraic immunity and the nonlinearity of our function is also higher than the ones mentioned, in addition to satisfying other cryptographic properties. Let $f_1 \in B_{n-2}$ be the Carlet–Feng function and $f_2(x) = f_1(Ax)$. For $n = 8, 10, 12, 14, 16$, in Table 3, we give the algebraic immunity and nonlinearity of $f$, compared to the same parameters for the function $TCT$ introduced by [35].

As it was also motivated in [38] for the HWBF, for the same number of variables, the algebraic immunity and nonlinearity of our construction may be lower than the ones of the Carlet-Feng function. However, since our functions can be implemented very efficiently in hardware (LFRSs are better suited for hardware), we can use more variables, thus increasing the cryptographic properties of the combiner.

Acknowledgements. The authors express their appreciation for the insightful and constructive comments of the referees and the editor, which improved the quality of the paper.

References


