On fractional correlation immunity of majority functions*

Chuan-Kun Wu
State Key Laboratory of Information Security, Institute of Software
Chinese Academy of Sciences, Beijing 100190, China
Email: ckwu@is.iscas.ac.cn

February 10, 2009

Abstract:
The correlation immunity is known as an important cryptographic measure of a Boolean function with respect to its resist against the correlation attack. This paper generalizes the concept of correlation immunity to be of a fractional value, called fractional correlation immunity, which is a fraction between 0 and 1, and correlation immune function is the extreme case when the fractional correlation immunity is 1. However when a function is not correlation immune in the traditional sense, it may also has a nonzero fractional correlation immunity, which also indicates the resistance of the function against correlation attack.
This paper first shows how this generalized concept of fractional correlation immunity is a reasonable measure on the resistance against the correlation attack, then studies the fractional correlation immunity of a special class of Boolean functions, i.e. majority functions, of which the subset of symmetric ones have been proved to have highest algebraic immunity. This paper shows that all the majority functions, including the symmetric ones and the non-symmetric ones, are not correlation immune. However their fractional correlation immunity approaches to 1 when the number of variable grows. This means that this class of functions also have good resistance against correlation attack, although they are not correlation immune in the traditional sense.

Key words: Cryptography, Majority function, Correlation immunity, Walsh transform.

1 Introduction

The development of cryptographic algorithms have experienced different attacks. As a result of the attacks, different measurement about the resistance against the corresponding attacks are proposed. When correlation attack [7] was treated as a threat, the concept of

*This work was supported by the NSFC no.60673068 and National 973 project no.2007CB807902.

correlation immunity is proposed [6] as a measurement about the resistance that a nonlinear combination function has against the correlation attack. Recently a new attack known as the algebraic attack is proved to be very effective to many stream ciphers as well as to some block ciphers. As a measurement of the resistance of a nonlinear function against the algebraic attack, another measurement known as algebraic immunity is proposed. The idea of algebraic attack is to find an annihilator of the targeting combining function. By doing so, the process of algebraic attack is to solve a system of nonlinear equations. When the algebraic degree of the annihilator is low, the computational complexity to solve such a system of nonlinear equations is also low. So the effectiveness of algebraic attack depends on whether one can find such an annihilator with low algebraic degree. On the other hand, when the combining function is of high algebraic immunity, the algebraic degree of any of its annihilators cannot be very low. Hence, a significant job for the designers is to find combining functions with highest possible algebraic immunity. It has been proved [5] that the order of the algebraic immunity of a Boolean function in n variables cannot exceed $\lceil \frac{n}{2} \rceil$. If a Boolean function has algebraic immunity of order $\lceil \frac{n}{2} \rceil$, then this function is said to have the highest algebraic immunity.

In 2004, Dalai [3] studied the majority functions to be a class of Boolean functions with highest algebraic immunity, and [4] further proves that the majority functions are the only symmetric Boolean functions in odd number of variables with maximum algebraic immunity. While algebraic immunity is an important cryptographic measurement, very often the best performance with one cryptographic measurement will sacrifice the performance with other cryptographic measurements. In this paper we study the correlation immunity of the generalized majority functions, which include the majority functions in odd number of variables and newly defined such functions in even number of variables.

2 The correlation immunity for nonlinear combining functions

The concept of correlation immunity was proposed by Siegenthaler in 1984 [6]. It is a security measure against the correlation attack (also known as divide-and-conquer attack) of nonlinear combiners [7]. Therefore we first briefly describe the correlation attack of nonlinear combiners, which gives the rationale of why correlation immunity is a reasonable security measure against the correlation attack. This helps us to introduce the concept of fractional correlation immunity in the next section.

2.1 Preliminaries about Boolean functions

Let $GF(2)$ be a finite field of two elements 0 and 1, and $GF^n(2)$ be an n-dimensional vector space over $GF(2)$. A mapping from $GF^n(2)$ into $GF(2)$ is called a Boolean function in n variables, denoted by $f(x)$, where $x = (x_1, x_2, ..., x_n)$ is the shorthand form of a vector in $GF^n(2)$. Define the number of 1's in the coordinates of vector x as the **Hamming weight**

of this vector, and is denoted as $W_H(x)$. If $W_H(x) = \frac{n}{2}$, i.e., there are equal number of 0's and 1's in the coordinates of x, then x is called balanced.

For a Boolean function $f(x)$ in n variables, when x goes through all the possible values (vectors) of $GF^n(2)$, then $f(x)$ will have 2^n corresponding outputs. The vector of all the outputs of $f(x)$ is called the **truth table** of $f(x)$, which has dimension 2^n. Of course x has to follow a particular order when going through all the possible values of $GF^n(2)$. If we treat a binary vector as the binary representation of an integer, then when the integer takes all the values from 0 to $2^n - 1$, then the corresponding vector goes through all the elements in $GF^n(2)$. Traditionally, we let the value of the binary representation of the integer to go from 0 incrementally to $2^n - 1$. If we collect all the vectors where $f(x)$ takes value 1, then the collection is called the support of $f(x)$, denoted as $supp(f) = \{x : f(x) = 1\}$. The number of 1's in the truth table of $f(x)$ is called the Hamming weight of $f(x)$ and is denoted as $W_H(f)$. It is easy to see that $W_H(f)$ is the number of elements in $supp(f)$.

2.2 The correlation attack of nonlinear combiners

Nonlinear combiner is a popular pseudo-random sequence generator for stream ciphers. The basic structure of nonlinear combiner in stream ciphers is as shown in figure 1.

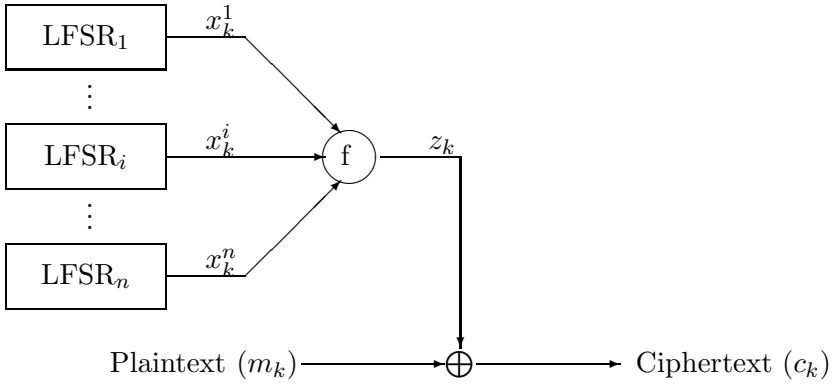

Figure 1: A nonlinear combiner of stream ciphers

The correlation attack proposed by Siegenthaler [7] makes use of the correlation information between the output sequence (z_k) of the nonlinear combiner and each input sequence (x_k^i) of the combining function $f(x)$, and to use the statistical analysis trying to recover the initial state as well as the feedback function of each LFSR$_i$ individually. This approach is also called *divide and conquer* attack, which significantly reduces the complexity than the brute force attack.

In the security analysis, it is always assumed that the structure of the generator is known, i.e. the lengths of each LFSR and the nonlinear combining function $f(x)$. The attack proposed in [7] does not assume the knowledge of the primitive feedback polynomial of each LFSR which is only of certain limited amount to search for.

Assume that all the LFSR's in the combiner of figure 1 are maximum length sequence

generators, i.e. each LFSR$_i$ of order r_i generates an m-sequence of period $p_i = 2^{r_i} - 1$, and there are R_i primitive polynomials of degree r_i (which is the number of different m-sequences of order r_i such that they are not equivalent by cyclic shift). Then under the brute force attack, the number of all the possible keys for the nonlinear combiner (different initial states and different feedback function of each of the LFSR have been taken into account) is

$$K = \prod_{i=1}^{n} R_i(2^{r_i} - 1).$$

With the correlation attack, information about each input sequence (x_k^i) can be extracted from the output sequence (z_k), and hence the attack can concentrate each of the individual LFSR sequences, and the number of trials in the worst case is reduced to approximately

$$K' = \sum_{i=1}^{n} R_i 2^{r_i}.$$

The correlation attack is a probabilistic attack which assumes some statistical properties of the combining function $f(x)$. Assume in the ideal case that each of the LFSR's in figure 1 produces a pseudo-random sequence with uniform probability distribution, i.e. $Prob(x_k^i = 0) = Prob(x_k^i = 1)$, and assume that $Prob(z_k = 0) = Prob(z_k = 1)$. Let

$$Prob(z_k = x_k^i) = q_i, \tag{1}$$

and assume the plaintext comes from a memoryless binary source, which satisfies

$$Prob(y_k = 0) = p_0 \tag{2}$$

Then it is easy to compute

$$\begin{aligned} Prob(c_k \oplus x_k^j = 0) &= Prob(z_k = x_k^j) \cdot Prob(y_k = 0) \\ &\quad + Prob(z_k \neq x_k^j) \cdot Prob(y_k = 1) \\ &= 1 - (p_0 + q_j) + 2p_0q_j \\ &= p_e \end{aligned} \tag{3}$$

When $j = 0$, let x_k^0 be an hypothetical random variable which are independent of any x_k^i $(i > 0)$ and with uniform probability distribution. Then compute the correlation of sequences c_k and x_k^j as

$$\alpha = \sum_{k=1}^{N}(1 - 2(c_k \oplus x_k^j)) = N - 2\sum_{k=1}^{N}(c_k \oplus x_k^j), \quad j \in \{0, 1, ..., n\} \tag{4}$$

By the central limit theorem, when N is sufficiently large, α approaches to a normal distribution (or Gaussian distribution). In an attack, attackers use hypothetical LFSR of length r_i which produce sequence (x_k^0) for the testing. By choosing a nonzero initial state and an arbitrary primitive polynomial as the feedback polynomial, compute the correlation α_0 between N bits of output of the hypothetical LFSR and N bits of the real ciphertext. Then there are two hypotheses to consider:

H_1**:** There are $N > r_i$ coincidences between the output of the hypothetical LFSR and LFSR_i, referring to the above cases, this is the case when α_0 is the correlation between z_k and x_k^i, $i \in \{1, 2, ..., n\}$.

H_0**:** There are $N > r_i$ disagreement between the output of the hypothetical LFSR and LFSR_i, referring to the above cases, this is the case when α_0 is the correlation between z_k and x_k^0.

In order to make a decision about the two hypotheses, a threshold value T is needed. When $\alpha_0 < T$, then accept the hypothesis H_0, and when $\alpha_0 \geq T$, accept H_1. Let the probability density function of the probabilistic variable α be $P_{\alpha|H_k}(x)$. If $q_i = \frac{1}{2}$ or $p_0 = \frac{1}{2}$, then by Eqn. (3), we have $p_e = \frac{1}{2}$, in this case no decision can be made, because in this case the probability distribution of α under the two hypotheses is the same. Here the discussed attack depends on the number of wrong decisions, i.e. the number of cases when $\alpha \geq T$. So we define a *false alarm probability* $P_f = Prob(\alpha \geq T|H_0)$. In order to determine an appropriate threshold T, we also need to consider the probability $P_m = Prob(\alpha < T|H_1)$. We have

$$P_f = \int_T^{\infty} P_{\alpha|H_0}(x)dx \tag{5}$$

$$P_m = \int_{-\infty}^{T} P_{\alpha|H_1}(x)dx \tag{6}$$

With the help of the function

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{y^2}{2}} dy \tag{7}$$

we can get the following expressions:

$$P_f = Q(|\frac{T}{\sqrt{N}}|) \tag{8}$$

$$P_m = Q(|\frac{N(2p_e - 1) - T}{2\sqrt{N}\sqrt{p_e(1 - p_e)}}|) \tag{9}$$

Denote by

$$\gamma_0 = \frac{N(2p_e - 1) - T}{2\sqrt{N}\sqrt{p_e(1 - p_e)}}, \tag{10}$$

then the expression of P_f and P_m can be written as

$$P_f = Q(|\sqrt{N}(2p_e - 1) - 2\gamma_0\sqrt{p_e(1 - p_e)}|), \tag{11}$$

$$P_m = Q(|\gamma_0|). \tag{12}$$

In order to attack the stream cipher model as in figure 1, the following process is to be taken: first to determine the probability q_i by $f(x)$, and to determine the probability p_0 according to the coding method of the plaintext, then compute p_e using Eqn. (3). For any chosen probability P_m, by Eqn. (12) it is known that γ_0 is a constant, and from Eqn. (11) it is known that the false alarm probability $P(\alpha \geq T|H_0)$ is a function of N. In order to

recover LFSR$_i$, choose an arbitrary primitive polynomial as its feedback polynomial and an arbitrary nonzero state as its initial state, and let it produce a sequence, then compute the correlation between this sequence and the ciphertext sequence. For any event with $\alpha \geq T$, H_0 is accepted, i.e., the LSFR$_i$ is supposed to have been recovered. However the probability of event $\alpha \geq T$ is P_f, and our decision may be wrong. So we need to test more ciphertexts for all the events $\alpha \geq T$. If for all the $2^{r_i} - 1$ different states, the decision is always to reject H_1, then change another primitive polynomial and to repeat the test. In the worst case we need to test for about $R_i 2^{r_i}$ times. The false alarm probability depends on the length of ciphertext N. Choose N_1 such that

$$P_f = \frac{1}{R_i 2^{r_i}} \tag{13}$$

then by Eqn. (11) we have

$$\frac{1}{R_i 2^{r_i}} = Q(|\sqrt{N_1}(2p_e - 1) - 2\gamma_0\sqrt{p_e(1-p_e)}|) \tag{14}$$

using the inequality

$$Q(x) < \frac{1}{2}2^{-\frac{x^2}{2}}, \quad x > 0 \tag{15}$$

we can get an upper bound of N_1:

$$N_1 < \left[\frac{\frac{1}{\sqrt{2}}\sqrt{\ln(R_i 2^{r_i - 1})} + \gamma_0\sqrt{p_e(1-p_e)}}{p_e - \frac{1}{2}}\right]^2. \tag{16}$$

The above is a brief description of the correlation attack which is mainly from [7]. The inclusion of the description is to help to understand how correlation immunity and consequently the fractional correlation immunity serves as a counter-measure against the correlation attack.

The upper bound in (16) gives the length of required ciphertext to enable the attack on the model in figure 1. If the length of the ciphertext is no less than this upper bound, then when performing such an attack, the number of tests can be minimized and when a decision is made, the probability of false alarm is minimized. More detailed description of the correlation attack can be found in [7].

In order to resist the correlation attack as described above, the combining function $f(x)$ needs to have some special properties. Siegenthaler [6] introduced the concept of correlation immunity of Boolean functions, and we will see how such functions can have resistance against the correlation attack. Then we will show how much resistance a fractionally correlation immune function would have against the correlation attack.

2.3 Correlation immunity as a counter-measure against the correlation attack

In order to resist the correlation attack as described above, Siegenthalar [6] proposed the concept of correlation functions.

Definition 1 *Let $f(x)$ be a Boolean function in n variables. Treat $(x_1, x_2, ..., x_n)$ as n random variables over $GF(2)$ that are independent and have uniform probability distribution, i.e. each x_i is equally likely to be 0 or 1. If for any $1 \leq i_1 < i_2 < \cdots < i_k \leq n$, the value of $f(x)$ is statistically independent of $(x_{i_1}, x_{i_2}, ..., x_{i_k})$, i.e., for any $(a_1, a_2, ..., a_k) \in GF^k(2)$ and any $c \in GF(2)$, we always have $Prob(f(x) = c|(x_{i_1}, x_{i_2}, ..., x_{i_k}) = (a_1, a_2, ..., a_k)) = Prob(f(x) = c)$, then $f(x)$ is said to be* **correlation immune** *of order k, or briefly k-order correlation immune. The maximum number of k such that $f(x)$ is k-order correlation immune is called the* correlation immunity *of $f(x)$, and is denoted as $CI(f) = k$.*

There are many different but equivalent definitions of correlation immunity. One of such alternatives is that, when a Boolean function $f(x)$ is correlation immune of order k, then its support $supp(f)$ has the property that, the vector formed from any $(i_1, i_2, \cdots, i_k)$ coordinates of $supp(f)$ will equally likely to be any vectors in $GF^k(2)$ when x goes through all the values in $supp(f)$. In particular, when $k = 1$, then the support of a correlation immune function has the property that, any coordinate of the vectors in $supp(f)$ has equal chances to be 0 or 1. It is trivial to verify that, if a Boolean function is correlation immune of order k, then for any $m < k$, this function is also correlation immune of order m.

Consider a simple case, assume that the combining function in the nonlinear combiner is correlation of order 1, i.e., for any x_i, the probability that $f(x)$ takes any value is not affected by a pre-fixed value of x_i. By Eqn. (3) we have that $p_e = \frac{1}{2}$. Taking it into Eqn. (16), we get an infinity upper bound of N_1, which means that the number of ciphertext to conduct such an attack may be infinity and hence not possible.

While a Boolean function of correlation immunity of order 1 being the combining function seems to resist the correlation attack, it is only to the case when consider the individual LFSR's. When a linear combination of a few of the LFSR's is considered, higher order correlation immunity is correspondingly required to resist the correlation attack.

3 The fractional correlation immunity as a counter-measure against the correlation attack

It is noted that the correlation immunity is a cryptographic measure about the resistance against correlation attack, there can be cases where although a combining function is not correlation immune, however the correlation attack still consumes large amount of computation due to the function being "almost" correlation immune. We hereby define a measure about how close a function is to being correlation immune. This only makes sense for the functions that are not correlation immune. Motivated by Eqn. (3) and Eqn. (4), let us first consider a simple case, i.e., the balancedness of the i-th coordinate of all the vectors in $supp(f)$. If it has a good balance, then $f(x)$ has small correlation with x_i. If it is balanced, then $f(x)$ has no correlation with x_i. If for all $i \in \{1, 2, ..., n\}$, $f(x)$ has no correlation with x_i, then $f(x)$ is correlation immune (of order at least 1). Not expecting the correlation immunity of $f(x)$, we define the relative correlation of $f(x)$ with x_i as the difference between

the number of 0's and that of 1's in the i-th coordinates of vectors x in $supp(f)$, i.e.,

$$\varepsilon^{(i)}(f) = |\sum_{x \in supp(f)} (-1)^{x_i}| = |W_H(f) - 2\sum_{x \in supp(f)} x_i|.$$

By this definition, it is easy to see that the idea of correlation immunity is to find the maximum value of these relative correlations. If the maximum value is 0, then $f(x)$ must be correlation immune (of order 1 or higher), otherwise $f(x)$ is not correlation immune. However, in the case $f(x)$ is not correlation immune, the value of $\varepsilon^{(i)}(f)$ varies which indicates the different degrees that $f(x)$ has correlation with x_i. The correlation of $f(x)$ with any variable is hence defined as

$$\varepsilon(f) = \max_{i \in \{1,2,\ldots,n\}} \varepsilon^{(i)}(f).$$

For this consideration, we define the **fractional correlation immunity** of $f(x)$ as

$$FCI(f) = 1 - \frac{\varepsilon(f)}{W_H(f)} = 1 - \frac{1}{W_H(f)} \max_{i \in \{1,2,\ldots,n\}} |W_H(f) - 2\sum_{x \in supp(f)} x_i| \qquad (17)$$

It is seen from Eqn. (17) that $0 \leq FCI(f) \leq 1$. When $FCI(f) = 1$, it means that $f(x)$ is correlation immune (of order 1). Another extreme case is when $FCI(f) = 0$, this means that there exists i such that $x_i = 0$ (or $x_i = 1$) always holds for all $x \in supp(f)$, which means that the correlation between $f(x)$ and this x_i is high (the highest possible case). In general, the fractional correlation immunity $FCI(f)$ is a fractional value between 0 and 1, instead of integral value as the traditional definition of correlation immunity.

Now we take a look at what the fractional correlation immunity has to do with the correlation attacks proposed by Siegenthaler. Let i be such an index satisfying that

$$FCI(f) = 1 - \frac{1}{W_H(f)}(|W_H(f) - 2\sum_{x \in supp(f)} x_i|) \stackrel{\Delta}{=} \varepsilon.$$

Then we have

$$Prob(x_i = 1 | f(x) = 1) = \frac{\sum_{x \in supp(f)} x_i}{|supp(f)|} = \frac{\sum_{x \in supp(f)} x_i}{W_H(f)}$$

and

$$\begin{aligned} Prob(x_i = 0 | f(x) = 0) &= \frac{|\overline{supp(f)}| - \sum_{x \in \overline{supp(f)}} x_i}{|\overline{supp(f)}|} \\ &= \frac{2^n - W_H(f) - (\sum_{x \in GF^n(2)} x_i - \sum_{x \in supp(f)} x_i)}{2^n - W_H(f)} \\ &= \frac{2^{n-1} - W_H(f) + \sum_{x \in supp(f)} x_i}{2^n - W_H(f)}. \end{aligned}$$

Hence by Eqn. (1) we have

$$q_i = Prob(f(x) = x_i)$$

$$
\begin{aligned}
&= Prob(f(x)=1)Prob(x_i=1|f(x)=1) + Prob(f(x)=0)Prob(x_i=0|f(x)=0) \\
&= \frac{W_H(f)}{2^n}\cdot\frac{\sum\limits_{x\in supp(f)} x_i}{W_H(f)} + \frac{2^n-W_H(f)}{2^n}\cdot\frac{2^{n-1}-W_H(f)+\sum\limits_{x\in supp(f)} x_i}{2^n-W_H(f)} \\
&= \frac{1}{2^n}(2^{n-1}-W_H(f)-2\sum_{x\in supp(f)} x_i) \\
&= \frac{1}{2}-\frac{W_H(f)}{2^n}(1-\varepsilon) \qquad (18)
\end{aligned}
$$

If ε is very close to 0, then q_i is very different from $\frac{1}{2}$. Particularly when $f(x)$ is balanced which is often practically required, then q_i is very close to 1 or 0, in which case, we have high confidence to have either $f(x)=x_i$ or $f(x)=x_i\oplus 1$. Consequently by Eqn. (3), we get that $p_e\approx p_0$ or $p_e\approx 1-p_0$. It is assumed that $p_0\neq\frac{1}{2}$, otherwise we would always have $p_e=\frac{1}{2}$ and hence the correlation attack does not work. It is also easy to verify that these are cases when $|p_e-\frac{1}{2}|$ reaches the maximum value, and by (16) we know that the minimum amount of data is needed to perform a correlation attack.

If ε is very close to 1, then q_i is very close to $\frac{1}{2}$, and by Eqn. (3), p_e is also very close to $\frac{1}{2}$, and consequently large amount of ciphertext is required to perform a correlation attack. Although such an attack is possible, however, when ε is so close to 1 that results in the bound of (16) to be too large to reach in practice, then the correlation attack becomes practically impossible.

To be more precise, taking Eqn. (18) into Eqn. (3) we have

$$p_e = 1-(p_0+q_i)-2p_0q_i = \frac{1}{2}+\frac{W_H(f)}{2^n}(1-\varepsilon)(\frac{1}{2}-p_0)$$

It is obvious that when ε is very close to 1, then p_e is also very close to $\frac{1}{2}$, and hence Eqn. (16) gives a very large upper bound.

Perhaps an upper bound of the size of text needed to conduct an attack is less convincing, because the actual number of text needed can be much smaller than the upper bound. To be more convincing, here we introduce a lower bound given in [8].

It is easy to prove that for the function $Q(x)$ defined in Eqn. (7), we have

$$Q(x) > \frac{1}{4}e^{-x^2}, \quad x>0$$

Taking into Eqn. (14) we have

$$
\begin{aligned}
N_1 &> \left(\frac{\sqrt{\ln(R_i2^{r_i})-2\ln 2}+2\gamma_0\sqrt{p_e(1-p_e)}}{2p_e-1}\right)^2 \\
&> \frac{\ln(R_i2^{r_i})-2\ln 2}{[W_H(f)(\frac{1}{2}-p_0)(1-\varepsilon)]^2} \qquad (19)
\end{aligned}
$$

This means that as long as the fractional correlation immunity of $f(x)$ is sufficiently close to 1, then p_e can be sufficiently close to $\frac{1}{2}$ and hence N_1 is sufficiently large, too large to be practically possible.

The concept of higher order fractional correlation immunity has similar motivation to that of higher order correlation immunity, i.e., it is to measure the probability of event $(f(x) = x_{i_1} \oplus x_{i_2} \oplus \cdots \oplus x_{i_k})$ and a corresponding modified correlation attack, for any possible $1 \leq i_1 < i_2 < \cdots < i_k \leq n$.

4 On the Walsh characterization of correlation immunity and fractional correlation immunity

Walsh transform has been a very useful tool in analyzing cryptographic properties of Boolean functions. Here we use Walsh transform to study the correlation immunity of majority functions.

Definition 2 *Let $f(x)$ be a Boolean function in n variables. The following function defined on the field of real numbers*

$$S_f(w) = \sum_{x=0}^{2^n-1} f(x)(-1)^{w \cdot x} \tag{20}$$

is called the **Walsh transform** *of $f(x)$, and the truthtable of $S_f(w)$ is called the Walsh spectrum of $f(x)$, where $w \cdot x = w_1x_1 \oplus w_2x_2 \oplus \cdots \oplus w_nx_n$ is the inner product of vectors w and x. For any $w \in GF^n(2)$, the value of $S_f(w)$ is called the Walsh spectrum of $f(x)$ on w.*

In the implementation of electronic circuits, it would be more convenient to use $\{-1,\ 1\}$ to represent the domain of binary functions than to use $\{0,\ 1\}$, and hence the following transform is used to map $\{0,\ 1\}$ to $\{-1,\ 1\}$:

$$F(x) = (-1)^{f(x)}.$$

By this transform, the Boolean function $f(x)$ is then mapped to function $F(x)$ on the domain $\{-1,\ 1\}$. The Walsh transform can also apply to $F(x)$ as:

$$S_F(w) = \sum_{x=0}^{2^n-1} F(x)(-1)^{w \cdot x} \tag{21}$$

Note that when the Boolean function $f(x)$ is also treated as a binary real valued-function, the Walsh transform remains the same. By this treatment, the two functions can be converted to each other:

$$F(x) = 1 - 2f(x)$$

Then, the Walsh transform of $F(x)$ can be converted from the Walsh transform of $f(x)$, i.e.

$$\begin{aligned} S_F(w) &= \sum_{x=0}^{2^n-1} (1 - 2f(x))(-1)^{w \cdot x} \\ &= \sum_{x=0}^{2^n-1} (-1)^{w \cdot x} - 2 \sum_{x=0}^{2^n-1} f(x)(-1)^{w \cdot x} \\ &= \begin{cases} 2^n - 2S_f(w) & \text{if } w = 0 \\ -2S_f(w) & \text{if } w \neq 0 \end{cases} \end{aligned}$$

On the other hand, the Walsh transform of $f(x)$ can be converted from the Walsh transform of $F(x)$:

$$S_f(w) = \begin{cases} 2^{n-1} - \frac{1}{2}S_F(w) & \text{if } w = 0 \\ -\frac{1}{2}S_F(w) & \text{if } w \neq 0 \end{cases}$$

Note that the Walsh transform on $F(x)$ is also known as the type II Walsh transform of $f(x)$, and is denoted as

$$S_{(f)}(w) = \sum_{x=0}^{2^n-1} (-1)^{f(x)+w\cdot x}.$$

And hence $S_f(w)$ is called the type I Walsh transform of $f(x)$. The following study will mainly use type II Walsh transform, and without confusion, we will simply call it Walsh transform, and its type can be distinguished by the notation.

There is a very good Walsh spectrum description about the correlation immunity of Boolean functions.

Lemma 1 (Xiao-Massey[9]) *A sufficient and necessary condition for $f(x) \in \mathcal{F}_n$ to be correlation immune of order k is that for any $w \in GF^n(2)$ with $1 \leq W_H(w) \leq k$, we have $S_{(f)}(w) = 0$.*

Although we used the type II Walsh spectrum to describe the correlation immunity in the above lemma, it is obvious that to use any type of Walsh spectrum will be the same, because for any nonzero w, we always have that $S_f(w) = 0$ if and only if $S_{(f)}(w) = 0$.

In order to compute the fractional correlation immunity of a Boolean function, motivated by lemma 1, we will seek a Walsh spectrum description. It is easy to deduce that the Walsh spectrum of $f(x)$ on e_i (where e_i is such a vector in $GF^n(2)$ that its i-th coordinate is 1 and 0 elsewhere) is

$$\begin{aligned} S_{(f)}(e_i) &= \sum_{x \in GF^n(2)} (-1)^{f(x)+w\cdot x} \\ &= -2 \sum_{x \in supp(f)} (-1)^{e_i \cdot x} \\ &= -2 \sum_{x \in supp(f)} (-1)^{x_i} \\ &= -2 \sum_{x \in supp(f)} (1 - 2x_i) \\ &= 4 \sum_{x \in supp(f)} x_i - 2W_H(f) \end{aligned}$$

Therefore by Eqn. (17) we have

$$FCI(f) = 1 - \frac{1}{2W_H(f)} \max_i |S_{(f)}(e_i)| \tag{22}$$

Now we can compute the fractional correlation immunity of the majority functions using Eqn. (22).

The above defined fractional correlation immunity corresponds to the traditional correlation immunity of order 1, i.e., the correlation of the output of the function with only one of the inputs is considered. Similar to the correlation immunity of order higher than 1, it is easy to extend the concept of fractional correlation immunity to the case when consider the correlation of the output of the function with a linear combination of its input variables. Note that the basic correlation attack considers $q_i = Prob(f(x) = x_i)$, in general case, we may consider a non-zero linear combination of the LFSR sequences, and the linear combination can be written as $w.x$, where $w \in GF^n(2)$ is the coefficient vector. Now the probability

$$q_w = Prob(f(x) = w \cdot x)$$

needs to be considered. When $w = e_i$, we have $w \cdot x = x_i$ which is the special case. However even in the general case, one cannot afford to count all the possible linear combinations in a practical attack. So we can restrict that there are at most k LFSR sequences involved in a linear combination, where k is a security parameter. So we need to consider all the linear combinations $w \cdot x$ with $1 \leq W_H(w) \leq k$. Similar to the analysis of how the fractional correlation immunity is related to the basic correlation attack, we can define the k-order fractional correlation immunity of $f(x)$ in n Variables as follows: First we define

$$\varepsilon^{(w)}(f) = |\sum_{x \in supp(f)} (-1)^{w \cdot x}|$$

to be the correlation between the output of $f(x)$ and the linear combination of its inputs with w as the coefficient vector of the linear combination, and define

$$\varepsilon_k(f) = \max_{w:\ 1 \leq W_H(w) \leq k} \varepsilon^{(w)}(f)$$

Then the **k-order fractional correlation immunity** of $f(x)$ is defined as

$$FCI_k(f) = 1 - \frac{\varepsilon_k(f)}{W_H(f)}$$

In order to compute the fractional correlation immunity of a Boolean function, motivated by lemma 1, we will seek a Walsh spectrum description. It is easy to deduce that the Walsh spectrum of $f(x)$ on e_i (as defined before, e_i is such a vector in $GF^n(2)$ that its i-th coordinate is 1 and 0 elsewhere) is

$$\begin{aligned} S_f(e_i) &= \sum_{x \in GF^n(2)} f(x)(-1)^{e_i \cdot x} \\ &= \sum_{x \in supp(f)} (-1)^{e_i \cdot x} \\ &= \sum_{x \in supp(f)} (-1)^{x_i} \\ &= \sum_{x \in supp(f)} (1 - 2x_i) \\ &= W_H(f) - 2 \sum_{x \in supp(f)} x_i \end{aligned}$$

Therefore we have

$$FCI(f) = 1 - \frac{1}{W_H(f)} \max_i |S_f(e_i)| \tag{23}$$

Given the relationship of the two types of Walsh spectrums, we have $S_f(e_i) = -\frac{1}{2}S_{(f)}(e_i)$, and hence the fractional correlation immunity can be represented as

$$FCI(f) = 1 - \frac{1}{2W_H(f)} \max_i |S_{(f)}(e_i)| \tag{24}$$

We will use the concept of fractional correlation immunity to study the majority functions in the next section, and will see that although the majority functions are not correlation immune in the traditional sense, but their fractional correlation immunity tends to approach to 1 with the increase of n, which means that they also have a good resistance against the correlation attack.

For the case of k-order fractional correlation immunity, with similar analysis as above, it is easy to deduce the following:

$$FCI_k(f) = 1 - \frac{\max\limits_{w:\ 1 \le W_H(w) \le k} |S_f(w)|}{W_H(f)}. \tag{25}$$

When $k = 1$, it becomes the factional correlation immunity as defined above, i.e., the specification of "1-order" is often omitted in the description.

Recall that if function $f(x)$ is correlation immune of order k, then it must be correlation immune of any order $m < k$ as well. In other words, if we write $CI_k(f)$ to be the k-order correlation immunity of $f(x)$, i.e., $CI_k(f) = 1$ means that function $f(x)$ is k-order correlation immune (note that here k may not be the largest correlation immunity of $f(x)$), and $CI_k(f) = 0$ means that $f(x)$ is not k-order correlation immune. Then $CI_k(f) = 1$ implies that $CI_m(f) = 1$ holds for any $m < k$. Alternatively we can write this as

$$CI_k(f) \le CI_m(f), \quad m < k$$

where $CI_i(f)$ only takes integral value 0 or 1.

Note from Eqn. (25) that the introduction of fractional correlation immunity also holds a similar inequality

$$FCI_k(f) \le FCI_m(f), \quad m < k \tag{26}$$

which generalizes the above inequality on correlation immunity to the fractional case. This can be understood as: if $f(x)$ is correlation immune of order k (k can be 0), then it must be correlation of order $k-1$. However perhaps it may not be correlation immune of order $k+1$, but its $(k+1)$-order correlation immunity is a fractional value, which can be very close to 1, where in the sense of traditional correlation immunity, this fractional value is interpreted as 0.

5 Majority functions and their Walsh spectrum characterization

Symmetric Boolean functions have many interesting properties [2], and those in odd number of variables have maximum algebraic immunity [4]. A special subset of symmetric Boolean functions are the majority functions, who seem to be the only symmetric functions that reach the maximum algebraic immunity for small number of variables [1].

In this paper, we will extend the concept of majority functions by introducing set majority functions. The concept of set majority function is not very new, but the term is never specifically named before. It should be noted that a large number of set majority functions in even number of variables are not symmtric. It is also noted that the study on majority functions so far has mainly been restricted to those symmetric ones, except [3] where the nonlinearity of balanced set majority functions in even number of variables are considered.

5.1 Definitions

Definition 3 *Let n be an odd number. The following defined Boolean function $f(x)$ in n variables is called a* **majority function***:*

$$f(x) = \begin{cases} 0, & \text{if } W_H(x) \leq \frac{n-1}{2}; \\ 1, & \text{if } W_H(x) \geq \frac{n+1}{2}. \end{cases} \tag{27}$$

The natural meaning of the above defined majority function is that, when the majority of the n-bit input has value 1, then the function outputs 1 which means a TRUE value, and when the majority of the input has value 0, the function outputs 0 which means FALSE.

The definition of majority function is very natural for the case when the input has odd number of coordinates, i.e. the number of inputs n of the function is an odd number. When n is even, there is no natural way of defining majority functions, as there are cases where the input has equal number of 0 values and 1 values. For this case, we generalize the concept of majority function as follows:

Definition 4 *Let n be an even number. Define $S = \{x \in GF^n(2) : W_H(x) = \frac{n}{2}\}$ and $A \subseteq S$. Then*

$$f_A(x) = \begin{cases} 0, & \text{if } W_H(x) < \frac{n}{2} \text{ or } x \in A; \\ 1, & \text{if } W_H(x) > \frac{n}{2} \text{ or } x \in (S \setminus A). \end{cases} \tag{28}$$

is called a **set majority function**.

Definition 4 generalizes the concept of majority function in odd number of variables to the case in even number of variables, and hence without confusion, the set majority function may simply be called the majority function and is simply denoted as $f(x)$ (without the subindex 'A'). The definition can be treated as universal (i.e. applies to odd and even number of variables) since when n is odd, the set S is an empty set and so is A.

There are two extreme cases of the set majority functions, that is when $A = S$ and when $A = \phi$ which is an empty set. When $A = S$, Eqn. (28) becomes

$$f_1(x) = \begin{cases} 0, & \text{if } W_H(x) \leq \frac{n}{2}; \\ 1, & \text{if } W_H(x) > \frac{n}{2}. \end{cases} \tag{29}$$

which is called a **strict majority function**. When $A = \phi$, Eqn. (28) becomes

$$f_0(x) = \begin{cases} 0, & \text{if } W_H(x) < \frac{n}{2}; \\ 1, & \text{if } W_H(x) \geq \frac{n}{2}. \end{cases} \tag{30}$$

which is called a **loose majority function**. The meaning of the above two extreme cases can be interpreted as follows: The strict majority function takes value 1 only when there are absolutely more 1 values than 0 values in the input, otherwise it takes value 0, including the case when the input has equal number of 0's and 1's; The loose majority function takes value 1 as long as the number of 1 value inputs is no less than that of 0 value inputs, including the case when their numbers are equal, and it takes 0 only when there are absolutely less 1's than 0's in the input. In general case, the set majority function takes value 1 when there are absolutely more 1's than 0's in the input, and it takes value 0 when there are absolutely less 1's than 0's in the input, and in the case when the input has equal number of 0's and 1's, it has to check if the input is from set A or $S \setminus A$. For the former case the function takes value 0 and otherwise it takes value 1.

For any given even number n, the strict majority function and the loose majority function are uniquely determined, just as the case of majority function defined for odd n. However, in general, the set majority function is not uniquely determined yet, as it depends on the set A.

It is noted that, except the strict majority function and the loose majority function, all the set majority functions in general are not symmetric.

Theorem 1 *When n is odd, the majority functions in n variables are all balanced; when n is even, the (set) majority functions in n variables are balanced if and only if $|A| = \frac{|S|}{2}$, where $|A|$ is the cardinality of set A.*

Proof: When n is odd, by the definition 3, the Hamming weight of the majority function $f(x)$ is $W_H(f) = \binom{n}{\frac{n+1}{2}} + \binom{n}{\frac{n+1}{2}+1} + \cdots + \binom{n}{n}$. Note that

$$\begin{aligned} 2^n &= \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\frac{n-1}{2}} \\ &\quad + \binom{n}{\frac{n+1}{2}} + \binom{n}{\frac{n+1}{2}+1} + \cdots + \binom{n}{n} \\ &= \binom{n}{n} + \binom{n}{n-1} + \cdots + \binom{n}{\frac{n+1}{2}} + \binom{n}{\frac{n+1}{2}} + \binom{n}{\frac{n+1}{2}+1} + \cdots + \binom{n}{n} \\ &= 2W_H(f) \end{aligned}$$

Hence we have $W_H(f) = 2^{n-1}$ which means that $f(x)$ is balanced.

When n is even, the Hamming weight of majority function $f_A(x)$ is $W_H(f_A) = \binom{n}{\frac{n}{2}+1} + \binom{n}{\frac{n}{2}+2} + \cdots + \binom{n}{n} + |S \setminus A|$. For convenience of writing, let $\Delta = \binom{n}{\frac{n}{2}+1} + \binom{n}{\frac{n}{2}+2} + \cdots + \binom{n}{n}$ and $A' = S \setminus A$. Then Δ can also be expressed as:

$$\Delta = \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\frac{n}{2}-1}.$$

Hence we have

$$\begin{aligned} 2^n &= \binom{n}{0} + \binom{n}{1} + \cdots + \binom{n}{\frac{n}{2}-1} \\ &\quad + \binom{n}{\frac{n}{2}} \\ &\quad + \binom{n}{\frac{n}{2}+1} + \cdots + \binom{n}{n} \\ &= \Delta + \binom{n}{\frac{n}{2}} + \Delta \end{aligned}$$

Note that $|A| + |A'| = |S| = \binom{n}{\frac{n}{2}}$, from the above we have

$$2^n = 2\Delta + |A| + |A'|.$$

So, $f_A(x)$ is balanced $\Longleftrightarrow W_H(f_A) = \Delta + |A'| = 2^{n-1} \Longleftrightarrow \Delta + |A| = 2^{n-1} \Longleftrightarrow |A| = |A'|$ $\Longleftrightarrow |A| = \frac{|S|}{2}$. □

It is seen from the above theorem that there is a strict restriction on the size of A when the majority function is required to be balanced. What is the number of such balanced functions for a given even n? Since A is any subset of S that has half of the elements in S, there can be $\binom{|S|}{|S|/2}$ choices of A. To distinguish this special case with the general case, we call this case as balanced majority functions, because this class of functions are all balanced.

Denote by $C(n)$ the number of balanced majority functions. Then $C(n) = \binom{0}{0} = 1$ for any odd value n. When n is even, it is easy to prove that

$$C(n) = \binom{\binom{n}{n/2}}{\binom{n}{n/2}/2}.$$

From table 1 it can be seen that the size of $C(n)$ increases very fast with the increase of n.

For the general case, by Stirling formula: $n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}}$, we can get an approximation:

$$\binom{n}{n/2} \approx \frac{2^{n+1}}{\sqrt{2n\pi e^{\frac{1}{4n}}}} \approx \frac{2^{n+1}}{\sqrt{2n\pi}}$$

and hence

$$C(n) = \binom{\binom{n}{n/2}}{\binom{n}{n/2}/2} \approx 2^{\frac{2^{n+1}}{\sqrt{2n\pi}} - \frac{n}{2} + \frac{1}{4}} n^{\frac{1}{4}} \pi^{-\frac{1}{4}} e^{\frac{1}{8n}}$$

which increases supper exponentially with the increase of n. In this sense, the generalized majority functions in even number of variables are more applicable in practice for their large number of supplies.

n	C(n)
2	2
4	20
6	184756
8	112186277816662845432
10	3.63×10^{74}
12	3.72×10^{276}
14	1.85×10^{1031}
16	1.26×10^{3872}
18	4.33×10^{14633}
20	2.32×10^{55614}

Table 1: The number of balanced majority functions in even number of variables

5.2 The Walsh spectrum characterization

Since the definition of majority functions differs much for the cases when n is odd and when n is even, our discussion will treat each of the cases respectively. Note that, when we write the XOR of two vectors such as $x \oplus s$, it means the bit-wise XOR of vectors x and s. In particular, $x \oplus \mathbf{1}$ means the complement of x, i.e., all the coordinates of x is taken the complement by XORing with 1.

5.2.1 When n is odd:

First we notice the following property of this class of functions:

Theorem 2 *When n is odd, a Boolean function $f(x)$ defined in definition 3 satisfies:*

$$f(x \oplus \mathbf{1}) = f(x) \oplus 1.$$

Proof: By definition 3, $f(x) = 0 \Longleftrightarrow W_H(x) \leq (n-1)/2 \Longleftrightarrow W_H(x \oplus \mathbf{1}) \geq (n+1)/2 \Longleftrightarrow f(x \oplus \mathbf{1}) = 1$. Similarly, $f(x) = 1 \Longleftrightarrow f(x \oplus \mathbf{1}) = 0$. □

Theorem 3 *Let $f(x)$ be a majority function in n variables, where n is an odd number, then the Walsh transform of $f(x)$ satisfies:*

$$S_{(f)}(w) = \begin{cases} 0, & \text{if } W_H(w) \text{ is even;} \\ 2 \sum\limits_{W_H(x) \leq \frac{n-1}{2}} (-1)^{w \cdot x}, & \text{if } W_H(w) \text{ is odd.} \end{cases} \tag{31}$$

Proof: Since $f(x)$ is a majority function in odd number of variables, we have

$$S_{(f)}(w) = \sum_{x \in GF^n(2)} (-1)^{f(x)+w \cdot x}$$

$$
\begin{aligned}
&= \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot x} - \sum_{W_H(x)\ge\frac{n+1}{2}} (-1)^{w\cdot x} \\
&= \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot x} - \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot(\mathbf{1}\oplus x)} \\
&= \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot x} - \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot\mathbf{1}+w\cdot x} \\
&= \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot x} - \sum_{W_H(x)\le\frac{n-1}{2}} (-1)^{W_H(w)+w\cdot x} \\
&= \begin{cases} 0, & \text{if } W_H(w) \text{ is even;} \\ 2\sum\limits_{W_H(x)\le\frac{n-1}{2}} (-1)^{w\cdot x}, & \text{if } W_H(w) \text{ is odd.} \end{cases}
\end{aligned}
$$

□

It is noted that when n is even, all the majority functions are symmetric, and in this case, a more convenient Walsh characterization is to treat input with different Hamming weight, and such a characterization can be found in [3].

5.2.2 When n is even:

Since the loose majority functions and the strict majority functions are both symmetric, their Walsh spectrum characterization can be done by treating the difference of the Hamming weight of the input, and such characterization can be found in [3] as well.

However in the general case, since set majority functions are not symmetric, Walsh spectrum characterization in terms of different Hamming weight of the inputs makes no sense, so we treat individual inputs. From definition 4 it is known that the majority function in even number of variables is not uniquely determined, it depends on the set A. Denote by $A' = S \setminus A = \{x : x \in S \text{ and } x \notin A\}$ to be the complement set of A with respect to S, and define

$$
\begin{aligned}
A_1 &= \{x : x \in GF^n(2) \text{ and } W_H(x) < \frac{n}{2}\} \\
A_2 &= \{x : x \in GF^n(2) \text{ and } W_H(x) > \frac{n}{2}\} \\
A_3 &= \{x : x \in A \setminus (A \cap \bar{A})\} \\
A_4 &= \{x : x \in A' \setminus (A' \cap \bar{A}')\} \\
A_5 &= \{x : x \in A \cap \bar{A}\} \\
A_6 &= \{x : x \in A' \cap \bar{A}'\}
\end{aligned}
$$

where $\bar{A} = \{x \oplus \mathbf{1} : \ x \in A\}$. Then it is easy to prove the following:

Lemma 2 *The above defined sets satisfy the following:*

1. $|A_1| = |A_2|, |A_3| = |A_4|$, *Furthermore, if* $|A| = \frac{|S|}{2}$, *then we also have* $|A_5| = |A_6|$.

2. $f(x)|_{A_1} = 0$, $f(x)|_{A_2} = 1$, $f(x)|_{A_3} = 0$, $f(x)|_{A_4} = 1$, $f(x)|_{A_5} = 0$, $f(x)|_{A_6} = 1$, *where* $f(x)|_A$ *represents the constraint function of* $f(x)$ *whose variable* x *can only take values from* A.

3. *Define a map* $\phi(x) = x \oplus \mathbf{1}$. *It maps every coordinate of* x *to its complement, and for a set* $B \subset GF^n(2)$, *we denote* $\phi(B) = \{y = \phi(x) : \ x \in B\}$. *Then we have* $\phi^2(x) = x$, $\phi^2(B) = B$, *and* $\phi(A_1) = A_2$, $\phi(A_3) = A_4$, $\phi(A_5) = A_5$, $\phi(A_6) = A_6$.

Based on lemma 2, we can formulate the Walsh transform of the majority functions in even number of variables. First we give

Lemma 3 *Let* $V \subset GF^n(2)$ *be a self-complement set, i.e.* $\bar{V} = \{x \oplus \mathbf{1} : \ x \in V\} = V$, *then for any odd Hamming weight vector* $w \in GF^n(2)$, *we have* $\sum\limits_{x \in V} (-1)^{w \cdot x} = 0$.

Proof: Denote $U = \sum\limits_{x \in V} (-1)^{w \cdot x}$, then

$$\begin{aligned} U &= \sum_{x \in V} (-1)^{w \cdot (x \oplus \mathbf{1})} \\ &= \sum_{x \in V} (-1)^{w \cdot x + W_H(w)} \\ &= (-1)^{W_H(w)} \sum_{x \in V} (-1)^{w \cdot x} \quad \text{(Since } W_H(w) \text{ is odd)} \\ &= -U. \end{aligned}$$

Hence $U = 0$. □

Theorem 4 *Let* $f_A(x)$ *be a majority function in* n *variables. Then the Walsh transform of* $f_A(x)$ *is:*

$$S_{(f_A)}(w) = \begin{cases} \sum\limits_{x \in A_5} (-1)^{w \cdot x} - \sum\limits_{x \in A_6} (-1)^{w \cdot x}, & \text{if } W_H(w) \text{ is even;} \\ 2 \sum\limits_{x \in A_3 \text{ or } W_H(x) < \frac{n}{2}} (-1)^{w \cdot x}, & \text{if } W_H(w) \text{ is odd.} \end{cases} \tag{32}$$

Proof: By the definition of $f_A(x)$ with respect to the sets A and S, and note that $S = A \cup A' = A_3 \cup A_5 \cup A_4 \cup A_6$ and $GF^n(2) = S \cup A_1 \cup A_2 = \bigcup_{i=1}^{6} A_i$, by lemma 2 and lemma 3 we have

$$\begin{aligned} S_{(f_A)}(w) &= \sum_{x \in GF^n(2)} (-1)^{f_A(x) + w \cdot x} \\ &= \sum_{x \in A_1} (-1)^{w \cdot x} - \sum_{x \in A_2} (-1)^{w \cdot x} + \sum_{x \in A_3} (-1)^{w \cdot x} - \sum_{x \in A_4} (-1)^{w \cdot x} \\ &\quad + \sum_{x \in A_5} (-1)^{w \cdot x} - \sum_{x \in A_6} (-1)^{w \cdot x} \\ &= \sum_{x \in A_1} (-1)^{w \cdot x} - \sum_{x \in A_1} (-1)^{w \cdot (x \oplus \mathbf{1})} + \sum_{x \in A_3} (-1)^{w \cdot x} - \sum_{x \in A_3} (-1)^{w \cdot (x \oplus \mathbf{1})} \end{aligned}$$

$$
\begin{aligned}
&+ \sum_{x\in A_5}(-1)^{w\cdot x} - \sum_{x\in A_6}(-1)^{w\cdot x} \\
= &\sum_{x\in A_1}(-1)^{w\cdot x} - (-1)^{W_H(w)}\sum_{x\in A_1}(-1)^{w\cdot x} \\
&+ \sum_{x\in A_3}(-1)^{w\cdot x} - (-1)^{W_H(w)}\sum_{x\in A_3}(-1)^{w\cdot x} \\
&+ \sum_{x\in A_5}(-1)^{w\cdot x} - \sum_{x\in A_6}(-1)^{w\cdot x} \\
= &\begin{cases} \sum\limits_{x\in A_5}(-1)^{w\cdot x} - \sum\limits_{x\in A_6}(-1)^{w\cdot x}, & \text{if } W_H(w) \text{ is even;} \\ 2\sum\limits_{x\in A_1\cup A_3}(-1)^{w\cdot x}, & \text{if } W_H(w) \text{ is odd} \end{cases} \\
= &\begin{cases} \sum\limits_{x\in A_5}(-1)^{w\cdot x} - \sum\limits_{x\in A_6}(-1)^{w\cdot x}, & \text{if } W_H(w) \text{ is even;} \\ 2\sum\limits_{x\in A_3 or W_H(x)<\frac{n}{2}}(-1)^{w\cdot x}, & \text{if } W_H(w) \text{ is odd} \end{cases}
\end{aligned}
$$

□

6 On the non-correlation immunity of majority functions

In order to check if the majority functions are correlation immune of any order at all, we might first look at whether they are correlation immune of order 1, for this purpose, by Lemma 1, we only need to verify their Walsh spectrum on a vector w with Hamming weight 1. Without loss of generality, let the vector e_i, as defined before, be such whose i-th coordinate is 1 and 0 elsewhere.

Regarding the correlation immunity of majority functions, we have the following conclusion.

Theorem 5 *No of the majority functions defined in definition 3 and definition 4 is correlation immune.*

Proof: Item 3 of Lemma 4 in [3] and Lemma 1 has shown that all symmetric majority functions are not correlation immune of order 1, this includes all the majority functions in odd number of variables, and the loose majority and the strict majority function in even number of variables. So we only need to show that the set majority functions in even number of variables in the general case are also not correlation immune of order 1.

When n is even, by theorem 4 we have

$$
S_{(f_A)}(e_i) = 2\sum_{x\in A_1\cup A_3}(-1)^{e_i\cdot x} = 2[\sum_{W_H(x)<\frac{n}{2}}(-1)^{x_i} + \sum_{x\in A_3}(-1)^{x_i}]
$$

Among all the n-dimensional vectors x with $W_H(x) < \frac{n}{2}$, the number of such vectors that also satisfy that the i-th coordinate is 1 (and the other $n-1$ coordinates can have $0 \sim \frac{n}{2}-1$ of 1's) is

$$
\binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{\frac{n}{2}-1},
$$

and the number of such vectors whose i-th coordinate is 0 (and the other $n-1$ coordinates can have $0 \sim \frac{n-1}{2}$ of 1's) is

$$\binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{\frac{n}{2}}.$$

Therefore

$$\begin{aligned}\sum_{W_H(x)<\frac{n}{2}} (-1)^{x_i} &= (\binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{\frac{n}{2}}) \\ &\quad -(\binom{n-1}{0} + \binom{n-1}{1} + \binom{n-1}{2} + \cdots + \binom{n-1}{\frac{n}{2}-1}) \\ &= \binom{n-1}{\frac{n}{2}}\end{aligned}$$

therefore we have

$$S_{(f_A)}(e_i) = 2[\binom{n-1}{\frac{n}{2}-1} + \sum_{x\in A_3} (-1)^{x_i}].$$

We show that the above is not always zero, i.e., if the above is zero for some i, then there must exist j such that $S_{(f_A)}(w_j) \neq 0$. Denote by $A_3^{i1} = \{x \in A_3 : \ x_i = 1\}$ and $A_3^{i0} = \{x \in A_3 : \ x_i = 0\}$, then $A_3 = A_3^{i0} \cup A_3^{i1}$.

Assume for some i, $S_{(f_A)}(e_i) = 0$, then $\sum_{x\in A_3} (-1)^{x_i} = 1 - \binom{n-1}{\frac{n}{2}-1}$. This means that $|A_3^{i1}| - |A_3^{i0}| = \binom{n-1}{\frac{n}{2}-1} - 1$. Note that when the i-th coordinate is fixed to be 1, the number of such vectors in S is $\binom{n-1}{\frac{n}{2}-1}$ (the other $n-1$ coordinates has $\frac{n}{2}-1$ of 1's). Since $|A_3^{i1}|$ cannot be larger than $\binom{n-1}{\frac{n}{2}-1}$, then there are only two cases: (1) $|A_3^{i1}| = \binom{n-1}{\frac{n}{2}-1}$ and $|A_3^{i0}| = 1$; or (2) $|A_3^{i1}| = \binom{n-1}{\frac{n}{2}-1} - 1$. We show that in both of the cases, there must exist a j such that $S_{(f_A)}(w_j) \neq 0$ and hence induces the conclusion of the theorem.

If case (1) is true, then the other $n-1$ coordinates (except i) of the vectors in A_3 have all the possible vectors of Hamming weight $\frac{n}{2}-1$. So for any $j \neq i$, there are $\binom{n-2}{\frac{n}{2}-1}$ elements in A_3^{i1} whose j-th coordinate is 0 (let the other $n-2$ coordinates take $\frac{n}{2}-1$ of 1's), and there are $\binom{n-2}{\frac{n}{2}-2}$ elements in A_3^{i1} whose j-th coordinate is 1 (let the other $n-2$ coordinates take $\frac{n}{2}-2$ of 1's). Since the j-th coordinate of the element in A_3^{i0} may be 0 or 1, we have

$$\sum_{x\in A_3} (-1)^{x_j} = \binom{n-2}{\frac{n}{2}-1} - \binom{n-2}{\frac{n}{2}-2} - c = \frac{(n-2)!}{\frac{n}{2}!(\frac{n}{2}-1)!} - c,$$

where $c \in \{0, 1\}$, which is larger than or equals to 0 when $n > 2$, and hence $S_{(f_A)}(w_j) > 0$.

If case (2) is true, then similarly other $n-1$ coordinates (except i) have all but one of the possible vectors of Hamming weight $\frac{n}{2}-1$. So for any $j \neq i$, there are $\binom{n-2}{\frac{n}{2}-1} - c_1$ elements in A_3^{i1} whose j-th coordinate is 0 (let the other $n-2$ coordinates take $\frac{n}{2}-1$ of 1's, taking away one such vector), and there are $\binom{n-2}{\frac{n}{2}-2} - c_2$ elements in A_3^{i1} whose j-th coordinate is 1 (let the other $n-2$ coordinates take $\frac{n}{2}-2$ of 1's, taking away one such vector). Hence we have

$$\sum_{x\in A_3} (-1)^{x_j} = \binom{n-2}{\frac{n}{2}-1} - c_1 - [\binom{n-2}{\frac{n}{2}-2} - c_2] = \frac{(n-2)!}{\frac{n}{2}!(\frac{n}{2}-1)!} + c_2 - c_1,$$

where $c_1, c_2 \in \{0,1\}$. As in case (1), when $n > 2$, it results in $S_{(f_A)}(w_j) > 0$. When $n = 2$, all the possible majority functions in 2 variables are: $f_1(x) = x_1$, $f_2(x) = x_2$, $f_3(x) = x_1x_2$ and $f_4(x) = x_1 \oplus x_2 \oplus x_1x_2$. It is easy to verify that no of these functions is correlation immune, and hence the conclusion of the theorem is true. □

7 On the fractional correlation immunity of majority functions

Given the Walsh characterization of the fractional correlation immunity and the Walsh spectrum characterization of majority functions, it is easy to link the two together to give an explicit description on the fractional correlation immunity of majority functions. Let $f(x)$ be the majority function in n variables being considered.

7.1 When n is odd

By definition 3 it is known that, $f(x) = 1$ if and only if $W_H(x) \geq \frac{n+1}{2}$. Similar to the discussion above, among the vectors with Hamming weight being larger than or equals to $\frac{n+1}{2}$, the number of such vectors where the i-th coordinate is 0 (and the rest $n-1$ coordinates can have $\frac{n+1}{2} \sim n-1$ of 1's) is $\binom{n-1}{n-1} + \binom{n-1}{n-2} + \cdots + \binom{n-1}{\frac{n+1}{2}}$, and the number of such vectors where the i-th coordinate is 1 (and the rest $n-1$ coordinates can have $\frac{n-1}{2} \sim n-1$ of 1's) is $\binom{n-1}{n-1} + \binom{n-1}{n-2} + \cdots + \binom{n-1}{\frac{n-1}{2}}$. Therefore

$$\begin{aligned} S_{(f)}(e_i) &= -2 \sum_{x \in supp(f)} (-1)^{x_i} \\ &= 2[\binom{n-1}{n-1} + \binom{n-1}{n-2} + \cdots + \binom{n-1}{\frac{n-1}{2}}] \\ &\quad -2[\binom{n-1}{n-1} + \binom{n-1}{n-2} + \cdots + \binom{n-1}{\frac{n+1}{2}}] \\ &= 2\binom{n-1}{\frac{n-1}{2}} \end{aligned}$$

Note that here the value of $S_{(f)}(e_i)$ is independent of i, hence by Eqn. (22) we have

$$\begin{aligned} FCI(f) &= 1 - \frac{1}{2W_H(f)} \max_i |S_{(f)}(e_i)| \\ &= 1 - \frac{1}{W_H(f)} . \binom{n-1}{\frac{n-1}{2}} \end{aligned} \tag{33}$$

By definition 3 we know that, when n is odd, the majority functions are balanced, i.e., $W_H(f) = 2^{n-1}$, and hence the above becomes

$$FCI(f) = 1 - \frac{1}{2^{n-1}} \binom{n-1}{\frac{n-1}{2}}.$$

By Stirling formula: $n! \approx \sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}}$, we further have

$$\binom{n}{\frac{n}{2}} = \frac{n!}{(\frac{n}{2})^2} \approx \frac{\sqrt{2\pi} n^{n+\frac{1}{2}} e^{-n+\frac{1}{12n}}}{2\pi(\frac{n}{2})^{n+1} e^{-n+\frac{1}{3n}}} = \frac{2^{n+1} e^{-\frac{1}{4n}}}{\sqrt{2n\pi}}$$

Then

$$\binom{n-1}{\frac{n-1}{2}} \approx \frac{2^n}{\sqrt{2\pi(n-1)}} e^{-\frac{1}{4(n-1)}}$$

So

$$1 - \frac{1}{2^{n-1}} \binom{n-1}{\frac{n-1}{2}} \approx 1 - \sqrt{\frac{2}{\pi(n-1)}} e^{-\frac{1}{4(n-1)}} \approx 1 - \sqrt{\frac{2}{\pi(n-1)}}.$$

Summarize the discussion above, we have

Theorem 6 *When n is odd, the fractional correlation immunity of the majority functions is*

$$FCI(f) = 1 - \frac{1}{2^{n-1}} \binom{n-1}{\frac{n-1}{2}} \approx 1 - \sqrt{\frac{2}{\pi(n-1)}}. \tag{34}$$

7.2 When n is even

By theorem 4 we have

$$S_{(f_A)}(e_i) = 2 \sum_{x \in A_3 or W_H(x) < \frac{n}{2}} (-1)^{e_i \cdot x} = 2[\sum_{x \in A_3} (-1)^{x_i} + \sum_{W_H(x) < \frac{n}{2}} (-1)^{x_i}].$$

It is easy to verify that, among the n-dimensional vectors of Hamming weight less than $\frac{n}{2}$, the number of such vectors where the i-th coordinate is 0 (and the rest $n-1$ coordinates can have $0 \sim (\frac{n}{2}-1)$ of 1's) is $\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\frac{n}{2}-1}$, and the number of such vectors where the i-th coordinate is 1 (and the rest $n-1$ coordinates can have $0 \sim \frac{n}{2}-2$ of 1's) is $\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\frac{n}{2}-2}$. Therefore

$$\begin{aligned} \sum_{W_H(x) < \frac{n}{2}} (-1)^{x_i} &= [\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\frac{n}{2}-1}] \\ &\quad -[\binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{\frac{n}{2}-2}] \\ &= \binom{n-1}{\frac{n}{2}-1}. \end{aligned}$$

Note from the definition that $|S| = \binom{n}{\frac{n}{2}}$ and A_3 cannot have more than half of the elements in S (otherwise A_3 would have at least a pair of complement vectors which contradicts with its definition), i.e. $|A_3| \leq \frac{|S|}{2} = \binom{n}{\frac{n}{2}}/2$. Since

$$-|A_3| \leq \sum_{x \in A_3} (-1)^{e_i \cdot x} \leq |A_3|,$$

so we get a lower bound of $S_{(f_A)}(e_i)$:

$$\begin{aligned} S_{(f_A)}(e_i) &\geq 2[-|A_3| - \binom{n-1}{\frac{n}{2}-1}] \\ &\geq 2[-\binom{n}{\frac{n}{2}}/2 - \binom{n-1}{\frac{n}{2}-1}] \\ &= -2\binom{n}{\frac{n}{2}} \end{aligned}$$

and an upper bound

$$\begin{aligned} S_{(f_A)}(e_i) &= 2[|A_3| - \binom{n-1}{\frac{n}{2}-1}] \\ &\leq 2[\binom{n}{\frac{n}{2}}/2 - \binom{n-1}{\frac{n}{2}-1}] \\ &= 0. \end{aligned}$$

By definition 4, we have

$$\begin{aligned} W_H(f_A) &= \binom{n}{\frac{n}{2}+1} + \binom{n}{\frac{n}{2}+2} + \cdots + \binom{n}{n} + |A_3| \\ &\geq \binom{n}{\frac{n}{2}+1} + \binom{n}{\frac{n}{2}+2} + \cdots + \binom{n}{n} \\ &= 2^{n-1} - \binom{n}{\frac{n}{2}}/2. \end{aligned}$$

Therefore, by Eqn. (22) we have

$$\begin{aligned} FCI(f_A) &= 1 - \frac{1}{2W_H(f_A)} \max_i |S_{(f_A)}(e_i)| \\ &\geq 1 - \frac{2}{2^n - \binom{n}{\frac{n}{2}}} . \binom{n}{\frac{n}{2}} \\ &\approx 1 - \frac{4}{\sqrt{2n\pi} - 2}. \end{aligned}$$

Summarize the discussion above we have

Theorem 7 *When n is even, then the fractional correlation immunity of any majority function $f_A(x)$ in n variables satisfies*

$$FCI(f_A) \geq 1 - \frac{2}{2^n - \binom{n}{\frac{n}{2}}} \binom{n}{\frac{n}{2}} \approx 1 - \frac{4}{\sqrt{2n\pi} - 2}. \tag{35}$$

Noticing that when n is odd, by theorem 6 we have

$$\lim_{n\to\infty} FCI(f) = 1,$$

and when n is even, by theorem 7 we have

$$\lim_{n\to\infty} FCI(f_A) = 1,$$

this yields the following conclusion.

Theorem 8 *The fractional correlation immunity of the majority functions defined in definition 3 and definition 4 approaches to 1 with the increase of the number of variables.*

Theorem 8 means that the majority functions are almost correlation immune, and they are more close to being correlation immune with the increase of n. This asymptotic property is called approaching correlation immunity.

8 Concluding remarks

This paper proposed a new security measure of cryptographic Boolean functions called fractional correlation immunity. When the fractional correlation immunity reaches value 1, it is the correlation immunity in the traditional sense. How the new measure is related to the resistance against correlation attack when such a function is used as the combining function in a nonlinear combiner is studied.

This paper also studies the fractional correlation immunity of majority functions. It is shown that for the correlation immunity in the traditional sense, no majority function is correlation immune. However the fractional correlation immunity of majority functions approaches to 1 with the number of variables grows. This means that the majority logic functions also have good resistance against correlation attack.

It is noted that the results in this paper can be generalized to the case of k-order fractional correlation immunity, where the analysis is unavoidably more complicated. It should be pointed out that the concept of fractional correlation immunity can also be used to study other classes of Boolean functions.

References

1. A. Braeken and B. Preneel: "On the Algebraic Immunity of Symmetric Boolean Functions", In *Progress in Cryptology - INDOCRYPT 2005*, LNCS 3797, Springer-Verlag 2005, pp.35-48.

2. A. Canteaut and M. Videau: "Symmetric Boolean functions", *IEEE Trans. on Information Theory*, Vol. IT-51, No.8, 2005, pp.2791-2811.

3. D. K. Dalai, S. Maitra and S. Sarkar: "Basic theory in construction of Boolean functions with maximum possible annihilator immunity", *Design Codes and Cryptography*, Vol.40, No.1, 2006, pp.41-58.

4. N. Li and W.-F. Qi: "Symmetric Boolean functions depending on an odd number of variables with maximum algebraic immunity", *IEEE Trans. on Information Theory*, Vol.IT-52, No.5, 2006, pp.2271-2273.

5. W. Meier, E Pasalic, and C Carlet: "Algebraic attacks and decomposition of Boolean

functions", In *Advances in Cryptology-EUROCRYPT 2004*, LNCS 3027, Springer-Verlag 2004, pp.474-491.

6. T.Siegenthaler: "Correlation-immunity of nonlinear combining functions for cryptographic applications", *IEEE Trans. on Information Theory,* Vol. IT-30, No.5, 1984, pp.776-780.

7. T. Siegenthaler: "Decrypting a Class of Stream Ciphers Using Ciphertext Only", *IEEE Trans. on Computers*, Vol.C-34, No.1, 1985, pp81-85.

8. C.K.Wu, X.Wang: "Balanced source coding and its applications in stream ciphers", *ACTA Electronica Sinica*, Vol.21, No.7, 1993, pp.23-32 (in Chinese).

9. G.Z.Xiao and J.L.massey: "A spectral characterization of correlation-immune combining functions", *IEEE Trans. on Information Theory.* Vol.IT-34, No.3, 1988, pp.569-571.