Research Article

Yu Zhou*, Jianyong Hu, Xudong Miao, Yu Han, and Fuzhong Zhang

On the confusion coefficient of Boolean functions

https://doi.org/10.1515/jmc-2021-0012
received April 02, 2021; accepted July 11, 2021

Abstract: The notion of the confusion coefficient is a property that attempts to characterize confusion property of cryptographic algorithms against differential power analysis. In this article, we establish a relationship between the confusion coefficient and the autocorrelation function for any Boolean function and give a tight upper bound and a tight lower bound on the confusion coefficient for any (balanced) Boolean function. We also deduce some deep relationships between the sum-of-squares of the confusion coefficient and other cryptographic indicators (the sum-of-squares indicator, hamming weight, algebraic immunity and correlation immunity), respectively. Moreover, we obtain some trade-offs among the sum-of-squares of the confusion coefficient, the signal-to-noise ratio and the redefined transparency order for a Boolean function.

Keywords: Boolean function, differential power analysis, confusion coefficient, signal-to-noise ratio, redefined transparency order

MSC 2020: 94C10, 94A60, 06E30

1 Introduction

Side-channel attacks (SCAs), which can use physical emanation from the device, are based on the fact that cryptographic algorithms are implemented on a physical device. Differential power analysis (DPA) is the most widely used attack method in SCA [1]. Since DPA was proposed, many algorithms have been attacked, like DES, LILI-128 [2]. In these attacks, S-boxes (or multiple output nonlinear Boolean functions, or named by (n, m)-functions) are prominent targets for these attacks. This is because they allow us to distinguish clearly between correct and incorrect hypotheses on key guesses [1,3–5]. Some links between indicators of (n, m)-functions and SCAs were established in a mathematical point of view, and these indicators included the signal-to-noise ratio (denoted by $\mathcal{SNR}$) [4], the transparency order (denoted by $\mathcal{TO}$) [1] or the redefined transparency order (denoted by $\mathcal{RTO}$) [6] and the confusion coefficient (denoted by $\mathcal{CC}$) [7].

In 2004, when Guilley et al. studied noise sources occurring during DPA and electrical simulation of the DPA, they introduced $\mathcal{SNR}$ for (n, m)-functions and gave some bounds on $\mathcal{SNR}$ for different S-boxes. Known results implied that the S-box could resist against DPA attack very well, if an S-box has a small $\mathcal{SNR}$. In refs [8,9], they obtained an upper bound on $\mathcal{SNR}$ of balanced (n, m)-functions, and some deep relationships between $\mathcal{SNR}$ of (n, m)-functions and three other cryptographic indicators (the maximum value of the absolute value of the Walsh transform, the sum-of-squares indicator and the nonlinearity of its coordinates), respectively, and they proved that $\mathcal{SNR}$ of (n, m)-functions is not affine invariant.

* **Corresponding author: Yu Zhou,** Science and Technology on Communication Security Laboratory, Chengdu 610041, China, e-mail: zhouyu.zhy@tom.com

Jianyong Hu, Xudong Miao, Yu Han, Fuzhong Zhang: Science and Technology on Communication Security Laboratory, Chengdu 610041, China

In 2005, $\mathcal{TO}$ was introduced for (n, m)-functions based on single-bit DPA and the Hamming distance model in ref. [1]. The lower bound on $\mathcal{TO}$ for an (n, m)-function was obtained. In the same year, a relationship between $\mathcal{TO}$ and the nonlinearity of an (n, m)-function was given in ref. [3]. Later, a fast method for computing $\mathcal{TO}$ was given in ref. [10]. In 2017, Chakraborty et al. found a redundancy in $\mathcal{TO}$, and put up $\mathcal{RTO}$ [6], and deduced a lower bound on $\mathcal{RTO}$ using Walsh spectrum only. The new indicator was used in multi-bit DPA attack and is better than the original definition in ref. [1]. From the two definitions, we know that $\mathcal{TO}$ and $\mathcal{RTO}$ are the same for an $(n, 1)$-function (i.e, a Boolean function). In 2019, some bounds on $\mathcal{TO}$ for only one Boolean function were obtained in ref. [11], including some tight upper bounds on $\mathcal{TO}$ for certain Boolean functions. Meanwhile, a relationship of $\mathcal{TO}$ between any Boolean function and its decomposition Boolean functions was given in ref. [12], and the distributions of $\mathcal{TO}$ for 4-variable and 5-variable balanced Boolean functions are obtained.

In 2012, $\mathcal{CC}$ was proposed when they studied the confusion property of cryptographic algorithms in ref. [7]. Some experimental results of DPA and CPA on both AES and DES verify the statistical model with high accuracy and demonstrate effectiveness of the algorithmic confusion analysis in ref. [13]. A novel heuristic technique to generate S-boxes with better values of the confusion coefficient in terms of improving their side-channel resistance is given in ref. [14]. Some new methods for generating almost optimal 8-bit S-boxes having good theoretical DPA metrics (including $\mathcal{SNR}$, $\mathcal{TO}$ and $\mathcal{CC}$, etc.) are obtained in ref. [15].

So far, many results about $\mathcal{SNR}$ and $\mathcal{TO}$ for (n, m)-functions are obtained in a mathematical point of view, but little attempt has been made to analyze $\mathcal{CC}$ for (n, m)-functions. How to characterize $\mathcal{CC}$ for (n, m)-functions? What is the bound on $\mathcal{CC}$ for (n, m)-functions? When further investigating the in-depth relationships between the $\mathcal{CC}$ and other cryptography indicators it still appears to be an important issue. In this article, we will investigate these questions for $(n, 1)$-functions (a Boolean function with n variables).

The organization of this article is as follows. In Section 2, the basic concepts and notions are presented. In Section 3, we deduce a equivalent characterization of the confusion coefficients. In Section 4, we give some relationships between the sum-of-squares of the confusion coefficients and other cryptographical indicators. In Section 5, some relationships among the sum-of-squares of the confusion coefficient, the RTO and the SNR are given. Finally, Section 6 concludes this article.

2 Preliminaries

Let $\mathbb{B}_n$ denote the set of n-variable Boolean functions. The support of a Boolean function $f \in \mathbb{B}_n$ is defined as $\mathrm{Supp}(f) = \{(x_1, \ldots, x_n) \in \mathbb{F}_2^n | f(x_1, \ldots, x_n) = 1\}$. The Hamming weight of f is denoted by $wt(f)$, that is, $wt(f) = |\mathrm{Supp}(f)|$. We say that a Boolean function f is balanced if its truth table contains equal numbers of ones and zeros, i.e., $wt(f) = 2^{n-1}$. A Boolean function $f \in \mathbb{B}_n$ is affine if the algebraic degree of f is at most one, and the set of all affine functions is denoted by $\mathbb{A}_n$. An affine function with constant term equalling to zero is called a linear function.

In this article, let $\mathbf{0}^n = \{0, 0, \ldots, 0\}$ be a zero vector.

Definition 2.1. Let $f \in \mathbb{B}_n$. The Walsh spectrum of f is defined as

$$\mathcal{F}(f \oplus \varphi_\alpha) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus \alpha x}, \ \alpha \in \mathbb{F}_2^n,$$

where $\varphi_\alpha = \alpha x = \alpha_1 x_1 \oplus \alpha_2 x_2 \oplus \cdots \oplus \alpha_n x_n$, $\alpha = (\alpha_1, \ldots, \alpha_n)$, $x = (x_1, \ldots, x_n)$.

Based on the Walsh spectrum, the nonlinearity of $f \in \mathbb{B}_n$ can be determined by

$$\mathcal{N}_f = 2^{n-1} - \frac{1}{2} \max_{\alpha \in \mathbb{F}_2^n} |\mathcal{F}(f \oplus \varphi_\alpha)|.$$

Let $f, g \in \mathbb{B}_n$. The cross-correlation function of f and g is defined as

$$\Delta_{f,g}(\alpha) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus g(x \oplus \alpha)}, \quad \alpha \in \mathbb{F}_2^n.$$

If $f = g$, then the autocorrelation function of f is defined as

$$\Delta_f(\alpha) = \sum_{x \in \mathbb{F}_2^n} (-1)^{f(x) \oplus f(x \oplus \alpha)}, \quad \alpha \in \mathbb{F}_2^n.$$

Definition 2.2. The two indicators (σ_f, Δ_f) are called the global avalanche characteristics of a Boolean function $f \in \mathbb{B}_n$ (*GAC* [16]):

$$\sigma_f = \sum_{\alpha \in \mathbb{F}_2^n} [\Delta_f(\alpha)]^2, \quad \Delta_f = \max_{\alpha \in \mathbb{F}_2^n, \alpha \neq \mathbf{0}^n} |\Delta_f(\alpha)|.$$

For any Boolean function $f \in \mathbb{B}_n$, the smaller Δ_f and σ_f, the better the *GAC*.

In 2012, Fei et al. proposed a statistical model for DPA that has taken characteristics of both the physical implementation and cryptographic algorithm into consideration and established a quantitative relation between the success rate of DPA and a cryptographic system. Finally, they point that the side-channel property of the cryptographic algorithm is extracted by a novel algorithmic confusion analysis and presented a definition of confusion coefficient for S-boxes [7].

Definition 2.3. [7] Let $k_i, k_j \in \mathbb{F}_2^n$ be two keys. The confusion coefficient (*CC*) κ over (k_i, k_j) is defined as:

$$\kappa = \kappa(k_i, k_j) = \Pr[(\psi|k_i) \neq (\psi|k_j)] = \frac{N_{(\psi|k_i) \neq (\psi|k_j)}}{N_t},$$

where N_t is the total number of values for the relevant ciphertext bits, and $N_{(\psi|k_i) \neq (\psi|k_j)}$ is the number of occurrences for which different key hypotheses k_i and k_j result in different ψ values.

Picek et al. [14] pointed that the low confusion coefficient values (also referred to as high collision values) can make SCAs harder, i.e., they may require an increase in number of traces to yield the correct key candidate.

Carlet et al. studied the intrinsic resiliency of S-boxes against SCAs, and further gave the concrete form of the confusion coefficient for a Boolean function $f \in \mathbb{B}_n$ (e.g., a coordinate function of the S-box) [17]:

$$\kappa(k, k^*) = \frac{1}{2^{n+2}} \sum_{t \in \mathbb{F}_2^n} [f(t \oplus k^*) - f(t \oplus k)]^2,$$

where $t \in \mathbb{F}_2^n$ is one known plaintext, $k^* \in \mathbb{F}_2^n$ is the correct key and $k \in \mathbb{F}_2^n$ is the key.

In other words, this equation gives a characterization relationship between the confusion coefficient and a Boolean function. In this article, we will study the confusion coefficient of a Boolean function based on this equation.

3 General bounds on CC of Boolean functions

In this section, we give one equivalent characterization of the confusion coefficient and give an upper and a lower bound on the confusion coefficient. Finally, we get some relationships between the confusion coefficient and the nonlinearity (propagation criterion).

Lemma 3.1. *Let $f \in \mathbb{B}_n$. Then*

$$\kappa(k, k^*) = \frac{1}{8} - \frac{1}{2^{n+3}} \Delta_f(d),$$

where $d = k^* \oplus k$, $k^*, k \in \mathbb{F}_2^n$.

Proof. According to Definition 2.3, we have

$$\begin{aligned}
\kappa(k, k^*) &= \frac{1}{2^{n+2}} \sum_{t\in\mathbb{F}_2^n} [f(t \oplus k^*) - f(t \oplus k)]^2 \\
&= \frac{1}{2^{n+2}} \sum_{t\in\mathbb{F}_2^n} \left[\frac{1}{2}\left(1 - (-1)^{f(t\oplus k^*)}\right) - \frac{1}{2}(1 - (-1)^{f(t\oplus k)}) \right]^2 \\
&= \frac{1}{2^{n+2}} \sum_{t\in\mathbb{F}_2^n} \frac{1}{4} \left[(-1)^{f(t\oplus k)} - (-1)^{f(t\oplus k^*)}\right]^2 \\
&= \frac{1}{2^{n+2}} \sum_{t\in\mathbb{F}_2^n} \frac{1}{2}\left[1 - (-1)^{f(t\oplus k^*)\oplus f(t\oplus k)}\right] \\
&= \frac{1}{2^{n+3}} \sum_{t\in\mathbb{F}_2^n} \left[1 - (-1)^{f(t\oplus k^*)\oplus f(t\oplus k)}\right] \\
&= \frac{1}{2^{n+3}} \sum_{t\in\mathbb{F}_2^n} 1 - \frac{1}{2^{n+3}} \sum_{t\in\mathbb{F}_2^n} (-1)^{f(t\oplus k^*)\oplus f(t\oplus k)} \\
&= \frac{1}{2^{n+3}} \star 2^n - \frac{1}{2^{n+3}} \sum_{t\in\mathbb{F}_2^n} (-1)^{f(t\oplus k\oplus d)\oplus f(t\oplus k)} \\
&= \frac{1}{2^3} - \frac{1}{2^{n+3}} \sum_{t=y\oplus k\in\mathbb{F}_2^n} (-1)^{f(y\oplus d)\oplus f(y)} \\
&= \frac{1}{8} - \frac{1}{2^{n+3}} \Delta_f(d).
\end{aligned}$$

□

From Lemma 3.1, we can obtain the distribution of the confusion coefficient of a Boolean function, if we calculate the distribution of autocorrelation function of this Boolean function. Here we give an example of 3-variable Boolean functions, since there are 30 different distributions of autocorrelation function [18], thus we can give the distributions of the confusion coefficient in Table 1 (the distribution of autocorrelation function for a Boolean function is denoted by D-of-A, and the distribution of confusion coefficient for one distribution of autocorrelation function is denoted by D-of-CC).

Based on Lemma 3.1, we can give an upper bound and a lower bound on the confusion coefficient for any Boolean functions.

Theorem 3.2. *Let $f \in \mathbb{B}_n$. For $k^*, k \in \mathbb{F}_2^n$, then*

$$0 \le \kappa(k, k^*) \le \frac{1}{4},$$

where the left equation holds if and only if $\Delta_f(k \oplus k^*) = 2^n$, and the right equation holds if and only if $\Delta_f(k \oplus k^*) = -2^n$.

Proof. From the autocorrelation function, we know $-2^n \le \Delta_f(\alpha) \le 2^n$ for any $\alpha \in \mathbb{F}_2^n$. Thus, we have

$$0 = \frac{1}{8} - \frac{1}{2^{n+3}} 2^n \le \kappa(k, k^*) \le \frac{1}{8} - \frac{1}{2^{n+3}}(-2^n) = \frac{1}{4}.$$

Furthermore, we have $\kappa(k, k^*) = 0$ if and only if $\Delta_f(k \oplus k^*) = 2^n$ and $\kappa(k, k^*) = \frac{1}{4}$ if and only if $\Delta_f(k \oplus k^*) = -2^n$. □

Particularly, for any Bent function $f \in \mathbb{B}_n$, we have $\kappa(k, k^*) = \frac{1}{8}$ for any $k \oplus k^* \neq \mathbf{0}^n$, and $\kappa(k, k^*) = 0$ for any $k \oplus k^* = \mathbf{0}^n$. For any linear Boolean function $f \in \mathbb{B}_n$, we have $\kappa(k, k^*) = 0$ or $\frac{1}{4}$ for any $k, k^* \in \mathbb{F}_2^n$, it is because that $|\Delta_f(\alpha)| = 2^n$ for any $\alpha \in \mathbb{F}_2^n$. For any constant Boolean function $f(x) \in \mathbb{F}_2^n$, that is, $f = 0$ or $f = 1$, we have $\kappa(k, k^*) = 0$.

Table 1: The distribution of CC for 3-variable Boolean functions

Class	D-of-A	Num	D-of-CC
1	(8, 8, 8, 8, 8, 8, 8, 8)	2	{0}
2	(8, 8, 0, 0, 0, 0, 0, 0)	8	{0, 0.125}
3	(8, 0, 8, 0, 0, 0, 0, 0)	8	{0, 0.125}
4	(8, 0, 0, 8, 0, 0, 0, 0)	8	{0, 0.125}
5	(8, 0, 0, 0, 8, 0, 0, 0)	8	{0, 0.125}
6	(8, 0, 00, 0, 8, 0, 0)	8	{0, 0.125}
7	(8, 0, 0, 0, 0, 0, 8, 0)	8	{0, 0.125}
8	(8, 0, 0, 0, 0, 0, 0, 8)	8	{0, 0.125}
9	(8, 4, 4, 4, 4, 4, 4, 4)	16	{0, 0.0625}
10	(8, 4, 4, 4, −4, −4, −4, −4)	16	{0, 0.0625, 0.1875}
11	(8, 4, −4, −4, 4, 4, −4, −4)	16	{0, 0.0625, 0.1875}
12	(8, −4, 4, −4, 4, −4, 4, −4)	16	{0, 0.0625, 0.1875}
13	(8, −4, −4, 4, −4, 4, 4, −4)	16	{0, 0.0625, 0.1875}
14	(8, −4, −4, 4, 4, −4, −4, 4)	16	{0, 0.0625, 0.1875}
15	(8, −4, 4, −4, −4, 4, −4, 4)	16	{0, 0.0625, 0.1875}
16	(8, 4, −4, −4, −4, −4, 4, 4)	16	{0, 0.0625, 0.1875}
17	(8, 0, 0, 0, 0, −8, 0, 0)	8	{0, 0.125, 0.25}
18	(8, 0, 0, 0, −8, 0, 0, 0)	8	{0, 0.125, 0.25}
19	(8, 0, 0, 0, 0, 0, −8, 0)	8	{0, 0.125, 0.25}
20	(8, 0, 0, 0, 0, 0, 0, −8)	8	{0, 0.125, 0.25}
21	(8, 0, 0, −8, 0, 0, 0, 0)	8	{0, 0.125, 0.25}
22	(8, 0, −8, 0, 0, 0, 0, 0)	8	{0, 0.125, 0.25}
23	(8, −8, 0, 0, 0, 0, 0, 0)	8	{0, 0.125, 0.25}
24	(8, 8, 8, 8, −8, −8, −8, −8)	2	{0, 0.25}
25	(8, 8, −8, −8, 8, 8, −8, −8)	2	{0, 0.25}
26	(8, 8, −8, −8, −8, −8, 8, 8)	2	{0, 0.25}
27	(8, −8, 8, −8, 8, −8, 8, −8)	2	{0, 0.25}
28	(8, −8, 8, −8, −8, 8, −8, 8)	2	{0, 0.25}
29	(8, −8, −8, 8, 8, −8, −8, 8)	2	{0, 0.25}
30	(8, −8, −8, 8, −8, 8, 8, −8)	2	{0, 0.25}

Lemma 3.3. *Let $f \in \mathbb{B}_n$, $k^*, k \in \mathbb{F}_2^n$ and $k \oplus k^* \neq \mathbf{0}^n$. Then*

$$\kappa(k, k^*) \geq \frac{1}{4} - \frac{[2^n - 2\mathcal{N}_f]^2}{2^{n+3}}.$$

Proof. From ref. [19], we have

$$\mathcal{N}_f \leq 2^{n-1} - \frac{\sqrt{2^n + \Delta_{\max}}}{2},$$

where $\Delta_{\max} = \max\{|\Delta_f(\alpha)| : \alpha \in \mathbb{F}_2^n, \alpha \neq \mathbf{0}^n\}$. And from Lemma 3.1, we have

$$2^n - 2^{n+3}\kappa(k, k^*) = \Delta_f(d) \leq \Delta_{\max} \leq (2^n - 2\mathcal{N}_f)^2 - 2^n.$$

Thus, this result is proved. □

From Lemma 3.3, we know that the confusion coefficients become larger, if $\mathcal{N}_f$ becomes larger. That is to say that the two indicators (the confusion coefficients and nonlinearity) cannot achieve the best for a given Boolean function at the same time. Moreover, for any balanced Boolean function $f \in \mathbb{B}_n$ we have $\kappa(k, k^*) \geq \frac{1}{8} - 2^{-n/2} - \frac{1}{2^{n-1}}$, it is because that $\mathcal{N}_f \leq 2^{n-1} - 2^{n/2-1} - 2$.

Lemma 3.4. *Let $f \in \mathbb{B}_n$ satisfy the propagation criterion with respect to all but a subset $R \subset \mathbb{F}_2^n$, $k^*, k \in \mathbb{F}_2^n$. Then*

$$\kappa(k, k^*) \geq \frac{1}{4} - \frac{1}{8}|R|.$$

Lemma 3.4 is easy to be proved by the following relation in ref. [19]:

$$\mathcal{N}_f \geq 2^{n-1} - 2^{n/2-1}|R|^{1/2}.$$

Theorem 3.2 gives the upper and the lower bounds on confusion coefficient. At the same time, we will also analyze the value of the sum of the confusion coefficients.

Theorem 3.5. *Let $f \in \mathbb{B}_n$. For a given $k^* \in \mathbb{F}_2^n$, we have*

$$\sum_{k \in \mathbb{F}_2^n} \kappa(k, k^*) = 2^{n-3} - \frac{(2^n - 2wt(f))^2}{2^{n+3}}.$$

Proof. For any Boolean function $f \in \mathbb{B}_n$, we have

$$\sum_{d \in \mathbb{F}_2^n} \Delta_f(d) = [2^n - 2wt(f)]^2.$$

Thus,

$$\begin{aligned}
\sum_{k \in \mathbb{F}_2^n} \kappa(k, k^*) &= \sum_{k \in \mathbb{F}_2^n} \left[\frac{1}{8} - \frac{1}{2^{n+3}} \Delta_f(k \oplus k^*) \right] \\
&= \frac{1}{8} \sum_{k \in \mathbb{F}_2^n} 1 - \frac{1}{2^{n+3}} \sum_{k \in \mathbb{F}_2^n} \Delta_f(k \oplus k^*) \\
&= 2^{n-3} - \frac{1}{2^{n+3}} \sum_{k \in \mathbb{F}_2^n} \Delta_f(k \oplus k^*) \\
&= 2^{n-3} - \frac{[2^n - 2wt(f)]^2}{2^{n+3}}.
\end{aligned}$$

□

Remark 3.6. From Theorem 3.5, we know that the value of the sum of the confusion coefficients for a given Boolean function $f \in \mathbb{B}_n$, if we give the hamming weight of this Boolean function. In particular, for a balanced Boolean function $f \in \mathbb{B}_n$, we have $\sum_{k \in \mathbb{F}_2^n} \kappa(k, k^*) = 2^{n-3}$ for a given k^*.

4 Bounds on the sum-of-squares of the confusion coefficient of one Boolean function

In order to establish some new relationships between the confusion coefficient and the traditional cryptography indicators (such as the sum-of-square indicator, hamming weight, algebraic immunity, correlation immunity, etc.), we need to use the sum-of-squares of the confusion coefficient. At first, we give the relationship between the sum-of-squares of confusion coefficient and the sum-of-square indicator for a Boolean function.

For the convenience of description and research in the following, for a given $k^* \in \mathbb{F}_2^n$ we denoted the sum-of-squares of the confusion coefficient for a Boolean function by

$$\mathcal{K}_f(k^*) = \sum_{k \in \mathbb{F}_2^n} \kappa^2(k, k^*).$$

Theorem 4.1. *Let $f \in \mathbb{B}_n$. For a given $k^* \in \mathbb{F}_2^n$, we have*

$$\mathcal{K}_f(k^*) = 2^{n-6} - \frac{[2^n - 2wt(f)]^2}{2^{n+5}} + \frac{\sigma_f}{2^{2n+6}}.$$

Proof. From Lemma 3.1, we have

$$\kappa^2(k, k^*) = \left[\frac{1}{8} - \frac{\Delta_f(k \oplus k^*)}{2^{n+3}}\right]^2 = \frac{1}{64} - \frac{\Delta_f(k \oplus k^*)}{2^{n+5}} + \frac{\Delta_f^2(k \oplus k^*)}{2^{2n+6}}.$$

Then

$$\begin{aligned}
\mathcal{K}_f(k^*) &= \sum_{k \in \mathbb{F}_2^n} \left[\frac{1}{64} - \frac{\Delta_f(k \oplus\ k^*)}{2^{n+5}} + \frac{\Delta_f^2(k \oplus\ k^*)}{2^{2n+6}}\right] \\
&= \sum_{k \in \mathbb{F}_2^n} \frac{1}{64} - \sum_{k \in \mathbb{F}_2^n} \frac{\Delta_f(k \oplus\ k^*)}{2^{n+5}} + \sum_{k \in \mathbb{F}_2^n} \frac{\Delta_f^2(k \oplus\ k^*)}{2^{2n+6}} \\
&= 2^{n-6} - \frac{1}{2^{n+5}} \sum_{k \in \mathbb{F}_2^n} \Delta_f(k \oplus\ k^*) + \frac{1}{2^{2n+6}} \sum_{k \in \mathbb{F}_2^n} \Delta_f^2(k \oplus\ k^*) \\
&= 2^{n-6} - \frac{[2^n - 2wt(f)]^2}{2^{n+5}} + \frac{\sigma_f}{2^{2n+6}}.
\end{aligned}$$

□

Based on Theorem 4.1, we can give the upper bound and the lower bound on the sum-of-squares of the confusion coefficient for a Boolean function.

Theorem 4.2. *Let $f \in \mathbb{B}_n$ be a nonconstant Boolean function. For a given $k^* \in \mathbb{F}_2^n$, we have*

$$2^{n-6} + 2^{-6} - \frac{[2^n - 2wt(f)]^2}{2^{n+5}} \le \mathcal{K}_f(k^*) \le 2^{n-5} - \frac{[2^n - 2wt(f)]^2}{2^{n+5}},$$

where the left equation holds if and only if f is a bent function, and the right equation holds if and only if f is a linear Boolean function.

Proof. Zhang and Zheng [16] obtained $2^{2n} \le \sigma_f \le 2^{3n}$ for any non-constant Boolean function $f \in \mathbb{B}_n$, where $\sigma = 2^{2n}$ if and only if f is a bent function, and $\sigma = 2^{3n}$ if and only if f is a linear function. Thus, this result is easy to be proved. □

If $f \in \mathbb{B}_n$ is a constant Boolean function, that is, $f \equiv 0$ or $f \equiv 1$. Then $\Delta_f(\alpha) = 2^n$ for any $\alpha \in \mathbb{F}_2^n$, we have $\kappa(k, k^*) = 0$ for any given $k^* \in \mathbb{F}_2^n$. Thus, $\mathcal{K}_f(k^*) = 0$.

For any r-order plateaued Boolean function [20], we have $\sigma_f = 2^{3n-r}$, and $wt(f) = 0$, or $2^{n-1} + 2^{n-r/2-1}$, or $2^{n-1} - 2^{n-r/2-1}$. Thus, $\mathcal{K}_f(k^*) = 2^{n-6} - 2^{n-r-6}$ or $2^{n-6} + 2^{n-r-6}$ for a given $k^* \in \mathbb{F}_2^n$.

For any balanced Boolean function $f \in \mathbb{B}_n (n \ge 3)$, $\sigma_f \ge 2^{2n} + 2^{n+3}$ [21]. By this result, we have Corollary 4.3.

Corollary 4.3. *Let $f \in \mathbb{B}_n$ be a balanced Boolean function ($n \ge 3$). For a given $k^* \in \mathbb{F}_2^n$, we have*

$$2^{n-6} + \frac{2^n + 8}{2^{n+6}} \le \mathcal{K}_f(k^*) \le 2^{n-5}.$$

For $n = 3$, we find 56 balanced Boolean functions reaching this lower bound $\mathcal{K}_f(k^*) = 2^{n-6} + \frac{2^n+8}{2^{n+6}} = 0.15625$ for a given $k^* \in \mathbb{F}_2^n$.

Based on Corollary 4.3, we have Table 2.

Recall the definition of algebraic immunity given in [22, Definition 73]. Let $f \in \mathbb{B}_n$, the minimum algebraic degree of nonzero annihilators of f or of $f \oplus 1$ (i.e., of nonzero multiples of $f \oplus 1$ or of f), is called the algebraic immunity of f and is denoted by $AI(f)$. Based on this definition, we have Corollary 4.4.

Table 2: The bounds on $\mathcal{K}_f(k^*)$ for balanced Boolean functions

n	Lower bound on $\mathcal{K}_f(k^*)$	Upper bound on $\mathcal{K}_f(k^*)$
3	0.15625	0.25
4	0.2734275	0.5
5	0.5195531	1
6	1.0175781	2
7	2.0166016	4
8	4.0161133	8
9	8.0158691	16
10	16.015747	32

Corollary 4.4. *Let $f \in \mathbb{B}_n$ be a balanced Boolean function and $AI(f) = k$. For a given $k^* \in \mathbb{F}_2^n$ we have*

$$\mathcal{K}_f(k^*) \le 2^{n-6} + \frac{\left[\sum_{i=k-1}^{n-k}\binom{n-1}{i}\right]^2}{2^{n+4}}.$$

Corollary 4.4 is easy to be proved by the following relation in ref. [23]:

$$\sigma_f \le 2^{n+2}\left[\sum_{i=k-1}^{n-k}\binom{n-1}{i}\right]^2.$$

And if a balanced Boolean function satisfies the propagation criterion, then we have Corollary 4.5 by using Theorem 3 in ref. [24] and Theorem 4.1.

Corollary 4.5. *Let $f \in \mathbb{B}_n$ be a balanced Boolean function ($n \ge 3$) and f satisfy the propagation criterion with respect to $A \subset \mathbb{F}_2^n$ and $t = \#A$, $k^* \in \mathbb{F}_2^n$. Then*

$$\mathcal{K}_f(k^*) \ge \begin{cases} 2^{n-6} + \dfrac{2^{2n} + 2^6(2^n - 1 - t)}{2^{2n+6}}, & 0 \le t \le 2^n - 2^{n-3} - 1,\ t(even), \\ 2^{n-6} + \dfrac{2^{2n} + 2^6(2^n + 2 - t)}{2^{2n+6}}, & 0 \le t \le 2^n - 2^{n-3} - 1,\ t(odd), \\ 2^{n-6} + 2^{-6} + \dfrac{1}{2^6(2^n - 1 - t)}, & 2^n - 2^{n-3} - 1 < t \le 2^n - 2. \end{cases}$$

Example 4.6. In Corollary 4.5, we can find some balanced Boolean functions reaching the lower bound.

(1) There are 10,080 4-variable balanced Boolean functions reaching $\mathcal{K}_f(k^*) = 2^{n-6} + \frac{2^{2n} + 2^6(2^n - 1 - t)}{2^{2n+6}} = 0.2734375$ for $t = 9$ and a given $k^* \in \mathbb{F}_2^4$, and 37,330,944 5-variable Boolean functions reaching $\mathcal{K}_f = 2^{n-6} + \frac{2^{2n} + 2^6(2^n - 1 - t)}{2^{2n+6}} = 0.53125$ for $t = 15$ and a given $k^* \in \mathbb{F}_2^5$;

(2) There are 1,920 4-variable balanced Boolean functions reaching $\mathcal{K}_f(k^*) = 2^{n-6} + \frac{2^{2n} + 2^6(2^n + 2 - t)}{2^{2n+6}} = 0.3046875$ for $t = 0$ and a given $k^* \in \mathbb{F}_2^4$;

(3) There are 27,776 5-variable balanced Boolean functions reaching $\mathcal{K}_f(k^*) = 2^{n-6} + 2^{-6} + \frac{1}{2^6(2^n - 1 - t)} = 0.53125$ for $t = 30$ and a given $k^* \in \mathbb{F}_2^5$.

Corollary 4.7. *Let $f \in \mathbb{B}_n$ be a balanced mth order correlation immune function ($1 \le m \le n - 2$). For a given $k^* \in \mathbb{F}_2^n$, we have*

$$\mathcal{K}_f(k^*) \ge 2^{n-6} + 2^{2m-n-2}.$$

Proof. From [25], we know $\sigma_f \geq 2^{n+2m+4}$ for any mth order correlation immune balanced function f. By Theorem 4.1, we have this result. □

Remark 4.8. Let $f \in \mathbb{B}_n$ and a given $k^* \in \mathbb{F}_2^n$. In Theorem 4.1, Corollary 4.4 and Corollary 4.7, we find the following facts.

(1) $\mathcal{K}_f(k^*)$ becomes bigger, if m becomes bigger.

(2) $\mathcal{K}_f(k^*)$ becomes smaller, if σ_f becomes smaller.

(3) $\mathcal{K}_f(k^*)$ becomes smaller, if $AI(f)$ becomes bigger.

As designers, we want m to be bigger, $AI(f)$ to be bigger, σ_f to be smaller and $\mathcal{K}_f(k^*)$ to be smaller for one Boolean function, thus these indicators cannot be the best at the same time. However, the best values of $AI(f)$, σ_f and $\mathcal{K}_f(k^*)$ may be achieved at the same time. These results can provide theoretical support for S-box design and evaluation.

5 Relationships among $\mathcal{CC}$, $\mathcal{SNR}$ and $\mathcal{RTO}$

In the section, we give three relationships among the confusion coefficient, the SNR and the RTO.

In 2004, Guilley et al. [4] presented the SNR of (n, m)-function in Definition 5.1.

Definition 5.1. [4] Let $F = (f_1, \ldots, f_m)$ be an (n, m)-function and $f_i \in \mathbb{B}_n$ $(1 \leq i \leq m)$. The $\mathcal{SNR}$ of F is defined by

$$\mathcal{SNR}(F) = \frac{m \cdot 2^{2n}}{\sqrt{\sum_{\alpha \in \mathbb{F}_2^n} [\sum_{i=1}^{m} \mathcal{F}(f_i \oplus \varphi_\alpha)]^4}}.$$

If $m = 1$ in Definition 5.1, then F is an n-variable Boolean function (let $F = f$). We have

$$\mathcal{SNR}(f) = \frac{2^{2n}}{\sqrt{\sum_{\alpha \in \mathbb{F}_2^n} [\mathcal{F}(f \oplus \varphi_\alpha)]^4}}.$$

In 2017, Chakraborty et al. [6] presented the RTO of (n, m)-function in Definition 5.2.

Definition 5.2. [6] Let $F = (f_1, \ldots, f_m)$ be a balanced (n, m)-function and $f_i \in \mathbb{B}_n$ $(1 \leq i \leq m)$. The $\mathcal{RTO}$ of F, based on the cross-correlation properties of F, is defined by:

$$\mathcal{RTO}(F) = \max_{\beta \in \mathbb{F}_2^m} \left\{ m - \frac{1}{2^{2n} - 2^n} \sum_{a \in \mathbb{F}_2^{n*}} \sum_{j=1}^{m} \left| \sum_{i=1}^{m} (-1)^{\beta_i \oplus \beta_j} \Delta_{f_i, f_j}(a) \right| \right\}.$$

If $m = 1$ in Definition 5.2, then F is an n-variable Boolean function (let $F = f$). We have

$$\mathcal{RTO}(f) = 1 - \frac{1}{2^{2n} - 2^n} \sum_{\alpha \in \mathbb{F}_2^{n*}} |\Delta_f(\alpha)|.$$

Based on Definition 5.1 and Theorem 4.1, we have Theorem 5.3.

Theorem 5.3. *Let $f \in \mathbb{B}_n$. For a given $k^* \in \mathbb{F}_2^n$, we have*

$$\mathcal{K}_f(k^*) = 2^{n-6} \left[1 + \frac{1}{\mathcal{SNR}^2(f)} \right] - \frac{[2^n - 2wt(f)]^2}{2^{n+5}}.$$

Proof. According to the definition of $\mathcal{SNR}$ for any Boolean function, we have

$$\sum_{\alpha\in\mathbb{F}_2^n}\mathcal{F}^4(f\oplus\varphi_\alpha)=\frac{2^{4n}}{\mathcal{SNR}^2(f)}.$$

And from [19] we know

$$\sigma_f=\frac{1}{2^n}\sum_{\alpha\in\mathbb{F}_2^n}\mathcal{F}^4(f\oplus\varphi_\alpha).$$

Thus, for any balanced Boolean function f, we have

$$\begin{aligned}\mathcal{K}_f(k^*)&=2^{n-6}-\frac{[2^n-2wt(f)]^2}{2^{n+5}}+\frac{\sigma_f}{2^{2n+6}}\\&=2^{n-6}-\frac{[2^n-2wt(f)]^2}{2^{n+5}}+\frac{\sum_{\alpha\in\mathbb{F}_2^n}\mathcal{F}^4(f\oplus\ \varphi_\alpha)}{2^{3n+6}}\\&=2^{n-6}-\frac{[2^n-2wt(f)]^2}{2^{n+5}}+\frac{2^{4n}}{2^{3n+6}\mathcal{SNR}^2(f)}\\&=2^{n-6}+\frac{2^{n-6}}{\mathcal{SNR}^2(f)}-\frac{[2^n-2wt(f)]^2}{2^{n+5}}.\end{aligned}$$

□

From Theorem 5.3, on one hand, we know that the sum-of-squares of the confusion coefficients becomes larger, if the $\mathcal{SNR}$ becomes smaller. That is to say that the two indicators (the sum-of-squares of the confusion coefficients and $\mathcal{SNR}$) cannot achieve the best for a given Boolean function at the same time. On the other hand, if we have a value of $\mathcal{SNR}$, then we can easily know the value of the sum-of-squares of the confusion coefficients for a given Boolean function. It can be seen from Theorem 5.3 that $\mathcal{SNR}$ and the sum-of-squares of the confusion coefficients are mutually determined for the same Boolean function.

Thus, for any balanced Boolean function, we have Corollary 5.4.

Corollary 5.4. *Let $f\in\mathbb{B}_n$ be a balanced Boolean function. For a given $k^*\in\mathbb{F}_2^n$, we have*

$$\mathcal{K}_f(k^*)=2^{n-6}\left[1+\frac{1}{\mathcal{SNR}^2(f)}\right].$$

From [9], we have $1\le\mathcal{SNR}(f)\le 2^{n/2}$ for any Boolean function $f\in\mathbb{B}_n$, thus we obtain the same result with Theorem 4.2 based on Corollary 5.4.

In Table 3, we find 18 different values of $\mathcal{K}_f(k^*)$ for all balanced 5-variable Boolean functions and the number of Boolean functions in every value, this classification is consistent with Table 2 in ref. [12].

Based on Definition 5.2 and Theorem 4.1, we have Theorem 5.5.

Theorem 5.5. *Let $f\in\mathbb{B}_n$ be a balanced Boolean function. For a given $k^*\in\mathbb{F}_2^n$, we have*

$$\mathcal{K}_f(k^*)\ge 2^{n-6}+\frac{[2^n-(2^n-1)\mathcal{RTO}(f)]^2}{2^{n+6}}.$$

Proof. According to the definition of $\mathcal{RTO}(f)$ for any Boolean function f, we have

$$\sum_{\alpha\in\mathbb{F}_2^{n*}}|\Delta_f(\alpha)|=[2^{2n}-2^n][1-\mathcal{RTO}(f)].$$

Table 3: The distribution of $\mathcal{K}_f(k^*)$ for 5-variable balanced Boolean functions

Class	$\mathcal{K}_f(k^*)$	Number	Percentage
1	1.000000	62	0.00001
2	0.794922	15872	0.00264
3	0.671875	59520	0.00990
4	0.666016	833280	0.13863
5	0.625000	8680	0.00144
6	0.607422	555520	0.09242
7	0.595703	9999360	1.66356
8	0.589844	8888320	1.47872
9	0.583984	1666560	0.27726
10	0.578125	1145760	0.19062
11	0.572076	6249600	1.03973
12	0.560547	73773056	12.27341
13	0.554687	90549760	15.06450
14	0.548828	66662400	11.09043
15	0.542969	133324800	22.18086
16	0.537110	166656000	27.72608
17	0.531250	39025280	6.49252
18	0.525391	1666560	0.27726

And from the definition of σ_f for any Boolean function f and the aforementioned equation, we know

$$\begin{aligned}\sigma_f &= \sum_{\alpha\in\mathbb{F}_2^n}\Delta_f^2(\alpha)\\ &= \frac{\sum_{\alpha\in\mathbb{F}_2^n}|\Delta_f(\alpha)|^2\sum_{\alpha\in\mathbb{F}_2^n}1^2}{2^n}\\ &\geq \frac{\left[\sum_{\alpha\in\mathbb{F}_2^n}|\Delta_f(\alpha)|\right]^2}{2^n}\\ &= \frac{\left[2^n+\sum_{\alpha\in\mathbb{F}_2^{n*}}|\Delta_f(\alpha)|\right]^2}{2^n}\\ &= \frac{[2^n+(2^{2n}-2^n)(1-\mathcal{RTO}(f))]^2}{2^n}.\end{aligned}$$

According to Theorem 4.1, we have

$$\begin{aligned}\mathcal{K}_f(k^*) &= 2^{n-6}-\frac{[2^n-2wt(f)]^2}{2^{n+5}}+\frac{\sigma_f}{2^{2n+6}}\\ &\geq 2^{n-6}-0+\frac{\frac{[2^n+(2^{2n}-2^n)(1-\mathcal{RTO}(f))]^2}{2^n}}{2^{2n+6}}\\ &= 2^{n-6}+\frac{[1+(2^n-1)(1-\mathcal{RTO}(f))]^2}{2^{n+6}}\\ &= 2^{n-6}+\frac{[2^n-(2^n-1)\mathcal{RTO}(f)]^2}{2^{n+6}}.\square\end{aligned}$$

Based on Corollary 5.4 and Theorem 5.5, we have Theorem 5.6.

Theorem 5.6. *Let $f\in\mathbb{B}_n$ and $wt(f)=2^{n-1}$. Then*

$$\mathcal{SNR}(f)[2^n-(2^n-1)\mathcal{RTO}(f)]\leq 2^n.$$

Through the confusion coefficient of any Boolean function, we establish a relationship between $\mathcal{SNR}$ and $\mathcal{RTO}$. For any (n, m)-function, a relationship between $\mathcal{SNR}$ and $\mathcal{RTO}$ was obtained in ref. [8]. If $m = 1$, Theorem 5.6 improves the result in ref. [8].

6 Conclusions

In this article, we give some bounds on the confusion coefficient for a Boolean function. And we also give some relationships between the confusion coefficient and other cryptographic properties. Moreover, some links among the confusion coefficient, the SNR and RTO are determined first. From the designer point of view, we hope to construct S-boxes with low the confusion coefficient, but a good S-box cannot make the confusion coefficient and some cryptographic indicators reaching the best at the same time, we hope that these results of Boolean functions will help us to construct good (n, m)-functions or S-box in the next step.

Acknowledgments: The authors wish to thank the anonymous referees for their valuable comments to improve the presentation of this article.

Funding information: This work was supported in part by the Sichuan Science and Technology Program (No. 2020JDJQ0076).

Conflict of interest: Authors state no conflict of interest.

References

[1] Prouff E. DPA attacks and S-Boxes. In: Handschuh H and Gilbert H, editors. Fast Software Encryption-FSE 2005, LNCS. vol. 3557. Berlin, Heidelberg: Springer; 2005. p. 424–42.

[2] Hoch J, Shamir A. Fault analysis of stream ciphers. In: Joye M and Quisquater JJ, editors. Chryptographic Hardware and Embedded Systems, CHES 2004, LNCS. vol. 3156. Springer-Verlag; 2004, p. 240–53.

[3] Carlet C. On highly nonlinear S-boxes and their inability to thwart DPA attacks. International Conference on Cryptology in India. Berlin Heidelberg: Springer-Verlag; 2005. p. 49–62.

[4] Guilley S, Hoogvorst P, Pacalet R. Differential power analysis model and some results. In: Quisquater JJ, Paradinas P, Deswarte Y, and El Kalam AA, editors. Smart Card Research and Advanced Applications VI. IFIP International Federation for Information Processing. vol. 153. Boston, MA: Springer; 2004 p. 127–42. https://doi.org/10.1007/1-4020-8147-2_9.

[5] Heuser A, Rioul O, Guilley S. A theoretical study of Kolmogorov-Simirnov distinguishers: side-channel analysis and secure design. In: Prouff E, editor. Lecture notes in computer science. 8622. Paris, France: Springer; 2014. p. 9–28.

[6] Chakraborty K, Sarkar S, Maitra S, Mazumdar B, Mukhopadhyay D, Prouff E. Redefining the transparency order. Design Code Cryptogr. 2017:82(1–2):95–115.

[7] Fei Y, Luo Q, Adam Ding A. A statistical model for DPA with novel algorithmic confusion analysis. In: Prouff E, and Schaumont P, editors. Cryptographic hardware and embedded systems - CHES 2012. Lecture notes in computer science. vol 7428. Berlin, Heidelberg: Springer; 2012. p. 233–50. https://doi.org/10.1007/978-3-642-33027-8_14.

[8] Zhou Y, Dong X, Wei Y, Zhang F. A note on the signal-to-noise ratio of (n, m)-functions. Adv Math Commun. 2020. https://doi.org/10.3934/amc.2020117.

[9] Zhou Y, Zhao W, Chen ZX, Wang WQ, Dun XN. On the signal-to-noise ratio for Boolean functions. IEICE Trans Fundamental. 2020;E103-A(12):1659–65.

[10] Fan L, Zhou Y, Feng D. A fast implementation of computing the transparency order of S-Boxes, Young Computer Scientists, 2008, ICYCS 2008. The 9th International Conference for IEEE. 2008. p. 206–11.

[11] Wang QC, Stănică P. Transparency order for Boolean functions: analysis and construction. Des Codes Cryptogr. 2019;87:2043–59. https://doi.org/10.1007/s10623-019-00604-1.

[12] Zhou Y, Dong XF, Wei YZ. On the transparency order relationship between one Boolean function and its decomposition functions. J Inf Secur Appl. 2021;58:102738. https://doi.org/10.1016/j.jisa.2020.102738.

[13] Fei Y, AdamDing A, Lao J, Zhang L. A statistics-based fundamental, model for side-channel attack analysis. Fundamental, model for side-channel attack analysis. IACR Cryptology ePrint Archive, Report 2014. https://eprint.iacr.org/2014/152.pdf.
[14] Picek S, Papagiannopoulos K, Ege B, Batinal L, Jakobovic D. Confused by confusion: systematic evaluation of DPA resistance of various S-boxes. In: Meier W and Mukhopadhyay D, editors. Progress in Cryptology - INDOCRYPT 2014. INDOCRYPT 2014. Lecture Notes in Computer Science. vol. 8885. Cham: Springer; 2014. p. 374–90. https://doi.org/10.1007/978-3-319-13039-2_22.
[15] de la Cruz Jiménez RA. On some methods for constructing almost optimal S-Boxes and their resilience against side-channel attacks. IACR Cryptology ePrint Archive, Report 2018. https://eprint.iacr.org/2018/618.pdf.
[16] Zhang XM, Zheng YL. GAC-the criterion for global avalanche characteristics of cryptographic functions. J Univers Comput Sci. 1995;1(5):316–33.
[17] Carlet C, de Cherisey E, Gulley S, Kavut S, Tang D. Intrinsic resiliency of S-boxes against side-channel attacks-best and worst scenarios. IEEE Trans Informa Forensic Secur. 2021;16:203–18.
[18] Zhou Y. On the distribution of autocorrelation value of balanced Boolean functions. Adv Math Commun. 2013;7(3):335–47.
[19] Zhang XM, Zheng YL. Auto-correlations and new bounds on the nonlinearity of Boolean functions. EUROCRYPT'96 Proceedings, LNCS. vol. 1070. Berlin, Heidelberg: Springer-Verlag; 1996. p. 294–306.
[20] Zheng YL, Zhang XM. On the plateaued functions. IEEE Trans Inform Theory. 2001;47(3):1215–33.
[21] Son JJ, Lim JI, Chee S, Sung SH. Global avalanche characteristics and nonlinearity of balanced Boolean functions. Informa Proc Letters. 1998;65:139–44.
[22] Carlet C. Boolean functions for cryptography and coding theory. New York: Cambridge University Press; 2020.
[23] Lovanov M. Tight bound between nonlinearity and algebraic immunity. IACR Cryptology ePrint Archive, Report 2005. https://eprint.iacr.org/2005/441.pdf.
[24] Sung SH, Chee S, Park C. Global avalanche characteristic and propagation criterion of balanced Boolean functions. Informa Proc Letters. 1999;69:21–24.
[25] Maitra S. Autocorrelation properties of correlation immune Boolean functions. Proceedings of the Second International Conference on Cryptology in India: Progress in Cryptology, LNCS. 2247. Berlin Heidelberg: Springer; 2001. p. 242–53.